Ca' Foscari
University
of Venice

Master's Degree Programme in
Data Analytics for Business and Society

Final Thesis

Gold, Bitcoin, and Macroeconomic Dynamics:

An Empirical Analysis of Safe Haven, Hedging, and Speculative Asset Behavior in Financial Markets

Supervisor
Diana Barro

Graduand
Dipta Roy
Matriculation Number 898395

Academic Year
2025 / 2026

Abstract

This thesis examines the behaviour of gold, Bitcoin, and a set of macroeconomic and financial variables between January 2008 and August 2025, a period that spans the Global Financial Crisis, the COVID-19 shock, the post-2022 inflation, and the emergence of digital assets. It asks whether gold prices can be predicted from their own history, which macroeconomic factors explain gold's movements, and whether Bitcoin behaves as a hedge, a safe haven, or a speculative asset relative to gold and the S&P 500.

These questions are addressed with three methods. An ARIMA model tests whether past gold prices forecast future ones. A Vector Error Correction Model, estimated at monthly frequency, relates gold to CPI, unemployment, the 10-year Treasury yield, the U.S. Dollar Index, and the S&P 500; the monthly frequency matches the release schedule of the macroeconomic series and avoids the distortions that forward-filling introduces into daily data. Finally, rolling measures, a GARCH(1,1), and a DCC-GARCH, combined with regime splits based on the official CBOE VIX, compare the volatility and equity correlation of Bitcoin, gold, and the S&P 500 across calm and high-stress periods.

Gold's short-run price movements are close to a random walk: one-day-ahead forecasts are accurate because tomorrow's price is usually near today's, but the model has little to say at longer horizons. In the monthly VECM, unemployment shocks have the largest contribution among the macro variables in the selected specification, accounting for roughly a third of gold's forecast error variance at twelve months and being associated with a positive response. That share, however, is concentrated in the 2020 pandemic: once 2020 is excluded it falls to about a tenth, so the finding we rely on is the ranking of unemployment above the other macro variables and the sign of its effect, not the precise magnitude. CPI shocks explain about 7 percent and carry a negative average sign, contrary to the view that gold reliably hedges inflation in the short run. The dollar and yield channels are weaker still. Gold is therefore more consistent with a conditional defensive asset against broad economic deterioration than with a simple inflation hedge. Bitcoin, by contrast, is far more volatile than gold and has grown more closely tied to the S&P 500 since 2020. The evidence supports treating gold as a conditional hedge and defensive asset — one that does not satisfy the strictest safe-haven definition on the most stressed days but remains far steadier than Bitcoin — and Bitcoin as a speculative one, a distinction that has sharpened in the post-2020 period.

Keywords: gold, Bitcoin, safe haven, hedge, speculative asset, ARIMA, VECM, GARCH, DCC-GARCH, macroeconomic uncertainty, official CBOE VIX.

Declaration on the Use of AI Tools

In accordance with the University guidelines on the use of artificial intelligence tools in final papers and degree theses, I declare that I used ChatGPT, developed by OpenAI, during the preparation and revision of this thesis.

These tools were used as writing-support and revision tools at different stages of the work. In particular, they were used to improve language clarity, academic tone, structure, and coherence; to reformulate selected explanations of the econometric methods; to improve the presentation of selected empirical findings; and to check the consistency of selected sections with the supervisor's feedback.

AI tools were not used as an independent data source. The empirical analysis, data processing, model implementation, numerical results, tables, figures, citations, and final interpretations were reviewed and checked by me. Statistical outputs and figures used in the thesis were produced through the empirical work described in the thesis, and not independently generated by AI.

I have checked the accuracy, reliability, and academic appropriateness of the AI-assisted content. I remain fully responsible for the content, originality, accuracy, and academic integrity of this thesis.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Background and Context

Gold has held a special position in financial markets for centuries. It carries no default risk, represents no claim on any issuer, and its supply grows only slowly through mining, so investors treat it as a defensive asset when uncertainty, inflation, or financial stress rises — a role visible through the 2008–2009 financial crisis, the 2020 COVID-19 shock, and the 2022–2023 inflation surge. Around the same time Bitcoin emerged with a capped supply, decentralized issuance, and independence from central banks, earning the label digital gold, a claim this thesis sets out to test. Gold's price does not move on its own — it responds to the U.S. dollar, Treasury yields, inflation, unemployment, equities, and now Bitcoin — so we study it inside a wider macro-financial system, asking what kind of asset it is under different conditions and whether Bitcoin can credibly stand in for it.

1.2 Problem Statement

Financial-market participants regularly describe gold as an inflation hedge or safe-haven asset, but they use these labels loosely. A true hedge should move in an uncorrelated or opposite direction to the asset being hedged. A true safe haven should preserve value specifically during periods of severe market stress. An asset can also serve as a diversifier without being a safe haven if it sometimes reduces portfolio risk but does not reliably protect investors during crises. Bitcoin complicates the debate because it shares some institutional features with gold, but its observed price behavior may put it closer to a speculative technology or risk asset.

Our empirical problem is therefore threefold. First, if gold prices behave like a random walk, then historical gold prices alone may have limited forecasting power. Second, if macroeconomic variables do influence gold, we need to know which variables matter in practice: inflation, the dollar, interest rates, stock-market risk, unemployment, or a combination of them. Third, if Bitcoin can substitute for gold, we must compare its volatility and correlation structure directly with those of gold and equities, especially during crisis periods.

1.3 Research Questions

- **RQ1.** Can the historical price of gold be used to forecast its future price?
- **RQ2.** How do macroeconomic variables (CPI, unemployment, and the 10-year Treasury yield) and financial indicators (the U.S. Dollar Index and the S&P 500) influence the dynamics of gold?
- **RQ3.** How do the volatility and correlation patterns of Bitcoin compare with those of gold and the S&P 500, and does Bitcoin behave as a hedge, safe haven, or speculative asset? We use the official CBOE VIX to split the sample into high-stress and low-stress regimes and compare each asset's behaviour across them.

1.4 Research Gap and Contribution

The literature reviewed in Chapter 2 already contains random-walk evidence for gold, cointegration studies of its macro drivers, and dynamic-correlation comparisons of gold and Bitcoin. Three gaps remain, one for each research question.

The first concerns which macro variable matters most for gold. The determinants literature concentrates on inflation, the dollar, and interest rates, while the bilateral evidence on unemployment (Thaver and Lopez, 2016) shows that it matters without ranking it against the others. We rank these drivers by their share of gold's forecast error variance in a single system, and the ordering inverts the usual one: unemployment, not inflation, has the largest macro contribution. That ranking is stable across orderings, the cointegrating rank, and crisis-excluding sub-samples; its magnitude is tied largely to the 2020 pandemic, so we claim the ordering, not the size.

The second gap is one of sample. The gold–Bitcoin comparisons we review (Kumar et al., 2025; Snene Manzli et al., 2025) end around the January 2024 approval of U.S. spot Bitcoin ETFs, just as Bitcoin's institutional era begins. Our sample runs through August 2025, and inside that window we condition on VIX quantiles and apply the same stress test to both assets within one GARCH framework.

The third gap concerns the symmetry of that test. The existing comparisons treat gold's safe-haven status as fixed and ask only whether Bitcoin matches it. We grant gold no such status: the same test is applied to both, and the result is uncomfortable for the benchmark, since gold's equity correlation rises rather than falls on the most stressed days. Both assets fail the strict definition, differing in degree. Because three drivers are released monthly, we also estimate the cointegration system at that frequency rather than forward-filling them into daily data, a point Chapters 2 and 5 develop.

1.5 Scope and Approach

We address each research question with a model suited to its empirical nature. RQ1 is a univariate forecasting problem, which we model with ARIMA. RQ2 is a multivariate problem with non-stationary variables that share long-run movement, so we use a Vector Error Correction Model at monthly frequency. RQ3 concerns volatility and time-varying dependence, which we handle with GARCH(1,1) and DCC-GARCH. We report the correlation analysis alongside the model results as a useful first-pass diagnostic. Each method isolates a different aspect of gold's role in the financial system.

CHAPTER 2: LITERATURE REVIEW

2.1 Gold as a Hedge and Safe Haven

The economic literature distinguishes carefully between a hedge, a diversifier, and a safe haven. Baur and Lucey (2010) define a hedge as an asset that moves uncorrelated or opposite to another asset on average, while a safe haven is an asset that moves uncorrelated or opposite to another asset, specifically during periods of market stress. A diversifier sits between these definitions: it has positive but imperfect correlation on average and may or may not protect investors during crises. These distinctions matter because the popular press uses the term 'safe haven' loosely, treating any negative correlation as evidence of safe-haven status. The empirical literature is more rigorous and explicitly tests the property.

Baur and McDermott (2010) extend this framework internationally and show that gold acts as a hedge against developed-market equities and as a safe haven against extreme negative shocks in major developed markets. Their evidence broadly supports gold's defensive role but also shows that the safe-haven property depends on the country, the time horizon, and the type of shock. Joy (2011) and Reboredo (2013) examine gold's behavior relative to the U.S. dollar and find that gold acts as both a hedge and a partial safe haven against dollar movements, with the strength of the relation varying across periods.

Recent evidence has kept the picture conditional. Using a quantile vector-autoregression connectedness framework across the COVID-19 pandemic, the Russia–Ukraine war, and the 2023 Silicon Valley Bank collapse, Snene Manzli and Jeribi (2024) find that gold and Bitcoin both behave as hedges under normal conditions and as safe havens during the three crises at the median quantile, but that gold is the more prominent safe haven and outperforms Bitcoin, especially during the war and the banking-sector stress. A more skeptical reading comes from Acikgoz (2025), who applies multifractal cross-correlation analysis to gold and Bitcoin against developed-market equities, emerging-market equities, bonds, commodities, energy, and other cryptocurrencies. He concludes that, in his setup, neither gold nor Bitcoin qualifies cleanly as a hedge or safe haven for financial markets, with the gold–bond pair offering the only notable reduction in risk. The two studies disagree, which matters for our purposes: the safe-haven label is sensitive to the assets compared, the crises examined, and the method used to test it, so it cannot be settled by a single statistic.

For this reason the thesis does not rest its conclusions on a single number. Because gold's hedge and safe-haven properties are conditional and method-dependent, we examine them with several complementary tools: a multivariate cointegration system for the macro-financial relationships in RQ2, and dynamic volatility and correlation models for the stress behaviour in RQ3. The remaining sections of this chapter set out the literature behind each of those tools.

2.2 Gold and Inflation

Whether gold reliably hedges inflation remains an unsettled question, with empirical studies reaching conflicting conclusions across different samples and methods. Shafiee and Topal (2010), in a study of the world gold market spanning roughly four decades, examine the relationship between the gold price and several key influencing variables, including oil prices and global inflation, as part of a model for forecasting long-term gold prices. Their analysis illustrates that the gold price is shaped by a set of macro-financial forces acting together rather than by inflation in isolation, which already suggests that the inflation channel cannot be read off a single pairwise relationship.

More recent work suggests the relationship depends on the prevailing inflation regime rather than holding on average. Valadkhani and O'Mahony (2024) study U.S. gold and silver returns with a Markov regime-switching vector autoregression on monthly data running from 1975 to 2023, allowing the estimated relationships to switch over time. They find that gold's response to inflation is regime dependent: gold returns react sharply and persistently to inflation in high-inflation regimes but only weakly in low-inflation ones. This, they argue, is why some earlier studies conclude that gold's inflation-hedging effectiveness has declined—by pooling long samples, those studies average over the prolonged high-inflation episodes in which the hedge actually operates and dilute their effect. They also report that gold and silver play complementary roles, with gold the more effective hedge when inflation is high and silver, given its wider industrial use, more responsive to the business cycle. An inflation-hedge relationship can therefore be real and still fail to appear in a full-sample estimate.

Cointegration-based studies reach a compatible conclusion by another route. Al-Nassar (2024), using a nonlinear ARDL model with structural breaks, finds gold and consumer prices to be cointegrated—so gold hedges inflation in the long run—while finding no reliable short-run hedge, the same long-run/short-run split that our own results reproduce. Nguyen (2024), applying an ARDL framework to U.S. data, similarly reports a strong long-term hedging role for gold and a positive gold–inflation relationship in high-inflation periods, together with a negative coefficient on the dollar index, which indicates that gold hedges currency depreciation at the same time. The long-run hedge is therefore robust across methods and samples, even where it is weak or absent in the short run—the horizon at which our monthly VECM operates.

This has a direct bearing on our own empirical design. A simple correlation between gold and CPI is not enough, because inflation rarely moves on its own: episodes of rising consumer prices are usually accompanied by shifts in monetary policy, real interest rates, and the exchange rate, and these accompanying channels can push gold in the opposite direction to the pure inflation effect. When a central bank responds to inflation by raising nominal rates faster than prices rise, real yields increase and the opportunity cost of holding gold—which pays no income—goes up, which tends to depress the gold price even as inflation climbs. A raw correlation cannot separate these forces, and it can even be spurious, since gold and the price level both tend to drift upward over long samples, so two series can appear strongly related without one driving the other. To isolate the inflation effect from these confounding channels, we therefore need a multivariate model that includes the other macro-financial variables alongside CPI—which is the job of the Vector Error Correction Model in RQ2.

2.3 Bitcoin and Digital Gold

The "digital gold" label rests on a set of institutional similarities between Bitcoin and gold: a supply capped at twenty-one million coins, an issuance schedule no central bank can expand, and operation outside the traditional banking system. The argument is that if scarcity, statelessness, and costly production are what give gold its defensive value, then Bitcoin should perform a similar role during market distress. The debate over whether it actually does emerged after Bitcoin's price reached significant levels around 2017 and intensified after the COVID-19 shock in 2020. Dyhrberg (2016) provides early evidence that Bitcoin shares some hedging properties with gold against the dollar. However, later work questions this conclusion. Bouri et al. (2017) find that Bitcoin's hedging and safe-haven properties are weaker than gold's, and they vary considerably across countries and over time. More broadly, the empirical literature on cryptocurrencies as financial assets has tended to find that they form a distinct asset class, less integrated with traditional safe-haven assets than the digital-gold narrative suggests.

The post-2020 evidence is even less favourable to the label, and the most recent work ties the shift directly to Bitcoin's institutionalization. Wu (2025) examines Bitcoin's correlation with the Nasdaq 100 and the S&P 500 over 2018–2025 using rolling-window correlations, static coefficients, and an event-study design, and documents a structural rise in Bitcoin's co-movement with equities following key institutional milestones—in particular the launch of U.S. spot Bitcoin exchange-traded funds and the inclusion of MicroStrategy in the Nasdaq 100—with the correlation peaking near 0.87 in 2024. The correlation is regime-dependent, rising during “risk-on” expansions and easing only briefly during acute “risk-off” episodes, and Wu reads this as Bitcoin moving from a relatively isolated alternative asset toward an integrated part of the mainstream risk-asset universe. This works against the digital-gold thesis: the institutional adoption expected to make Bitcoin more legitimate and more gold-like has instead made it trade more like a technology stock.

The crisis-period comparisons agree. Snene Manzli and Jeribi (2024), evaluating both assets against G7 equities through the COVID-19, Ukraine, and Silicon Valley Bank episodes, find gold to be the more dependable safe haven and Bitcoin the weaker one, while Acikgoz (2025) withholds clean safe-haven status from both yet still finds Bitcoin the more speculative of the two on volatility-based criteria. Recent work therefore supports a qualified conclusion: digital gold was a defensible hypothesis when Bitcoin was a niche, lightly integrated asset, but since 2020 the data show it behaving as a high-volatility risk asset whose correlation with equities has grown rather than shrunk. This is what RQ3 sets out to test directly.

2.4 Volatility Indices and Market Stress

We measure market stress using the CBOE Volatility Index (VIX), widely known as the market's "fear gauge" because it rises when investors anticipate turbulence ahead. Whaley (2000) first characterised it in these terms, and Whaley (2009) sets out the index's origins and what it is designed to capture. What makes the VIX well suited to our purpose is that it is forward-looking and option-implied: it is extracted from the prices of S&P 500 index options and represents the market's expectation of volatility over the next thirty days, so it reacts to anticipated stress before that stress shows up in realized returns—unlike a backward-looking rolling volatility computed from past prices. Sarwar (2012) examines how the VIX and the U.S. stock market move in relation to one another and finds a strong inverse link: when the VIX rises, equity returns (gains or losses one makes from holding stocks or shares of ownership in companies) tend to fall, and vice versa. The relationship is also asymmetric, reacting far more sharply to negative market shocks than to positive ones, which is why the VIX functions as a measure of fear rather than of optimism. What matters most for this thesis is that the strength of this link is not constant: it intensifies during turbulent periods, when the VIX is already high and volatile. Sarwar and Khan (2019) extend the analysis beyond the United States and find that surges in U.S. market fear spill over into emerging markets, explaining a substantial share (22–42%) of the variation in their returns. The VIX therefore reflects a form of market stress that is transmitted internationally, not confined to U.S. equities.

Recent work also uses the VIX to tell defensive assets apart from speculative ones. Sokhanvar and Hammoudeh (2024) study how gold, the U.S. dollar, and Bitcoin respond to VIX movements around U.S. yield-curve inversions, using static and dynamic cross-quantilogram methods. Gold and the dollar respond positively to increases in the VIX, the behaviour expected of a safe haven, whereas Bitcoin responds negatively, the behaviour of a risk asset that is sold when fear rises. They also find that a VIX shock has an immediate effect that then fades over the following days, and that yield-curve inversions change the strength of these links, with the VIX–dollar relationship mattering more during turbulent post-inversion phases. Both findings support our choice to treat the VIX as a forward-looking stress signal and to study its effect through regime-conditional methods rather than a single static correlation.

This shapes how we treat the VIX in RQ3. Because the link between fear and market losses is strongest exactly when fear is already elevated, relying on the average relationship across the full sample would be misleading. We therefore divide the sample into calm and high-stress periods and compare how gold and Bitcoin behave in each. It is the comparison between stressed and calm conditions, not average behaviour, that tests safe-haven status.

2.5 Forecasting Gold Prices and the Random-Walk Problem

In our first research question, we ask whether historical gold prices can forecast future gold prices. Box and Jenkins (1976) is the standard reference for ARIMA modeling. Financial prices often behave like a random walk, meaning that the best short-term forecast is the most recent observed price. In such cases, a model may produce low one-step-ahead forecast errors without truly discovering deep predictability.

Recent applied work supports treating gold as close to a random walk. Pflaumer (2024) models the gold price as a geometric random walk within the ARIMA framework over the period from 1971 to 2024 and shows that the drift term, rather than any rich autoregressive structure, is what drives the long-run projection. That is the outcome one would expect from a series that carries momentum in its level but holds little exploitable structure in its day-to-day increments—which is the situation we confirm for our own sample in Chapter 6. It also explains why a parsimonious specification can forecast gold's level well at short horizons while telling us almost nothing about its direction further out.

This pattern is what the efficient-market view of a deep, liquid market predicts: if any simple rule for profiting from gold's own past prices existed, traders would exploit it until it disappeared, leaving the day-to-day increments close to unforecastable noise.

The recent machine-learning literature fits this picture. Ben Jabeur et al. (2024), comparing six machine-learning models and finding the XGBoost algorithm the most accurate, obtain their forecasting gains not from gold's price history alone but from a broad set of external predictors—macroeconomic, financial, and market variables—and use Shapley (SHAP) values to show which of those inputs the model actually relies on. This matches the random-walk evidence: gold's own past carries little exploitable structure, so genuine forecasting power has to come from outside the price series. That is the motivation for moving from the univariate ARIMA of RQ1 to the multivariate system of RQ2.

ARIMA models are well-suited to this question because they examine whether past values and past errors carry useful information for forecasting future values. If gold is non-stationary in levels but becomes stationary after first differencing, then an ARIMA model with one order of differencing fits the series. The autocorrelation and partial-autocorrelation functions of the differenced series, together with information criteria such as the AICc, guide the choice of the autoregressive and moving-average orders, and the fitted residuals are then checked for any leftover structure. In our results, we retain the ARIMA(0,1,0) model as the central benchmark because it directly represents the random-walk hypothesis.

2.6 Volatility Modeling and Dynamic Correlation

Volatility — how sharply an asset's price moves up and down — is central to this thesis, because we cannot judge whether an asset is defensive or speculative from its average return alone. Two assets can deliver the same average return while one barely moves and the other swings violently; the second is far riskier, and that risk is invisible if we look only at averages. A further complication is that volatility is not steady over time. Financial markets display volatility clustering: turbulent days tend to be followed by more turbulent days, and calm by more calm, so periods of high and low risk arrive in streaks rather than evenly. GARCH models were developed to capture exactly this, tracking how an asset's risk rises and falls over time instead of assuming it stays fixed. Engle (2002) extends this idea from a single asset to the relationship between two assets: his Dynamic Conditional Correlation (DCC) model allows the correlation between assets to shift over time rather than holding it constant. This matters for safe-haven analysis because what defines a safe haven is how an asset behaves during periods of market stress, not how it behaves on average. An asset might move independently of equities in calm times yet fall alongside them in a crisis — a distinction a time-varying correlation can show but an average cannot.

In practice this thesis uses the workhorse GARCH(1,1) specification, in which today's conditional variance depends on yesterday's squared return shock and yesterday's conditional variance. We estimate it with a Student-t error distribution to accommodate the fat tails of daily returns—extreme moves occur more often than a normal bell curve would imply—and it yields a single volatility series per asset on a common scale, which is the most direct way to compare gold, the S&P 500, and Bitcoin. The same machinery produces a news-impact curve, which shows how a given return shock feeds into future volatility and makes the gap between a high-risk and a low-risk asset visible at every shock size. The DCC-GARCH then extends this from one asset to several in a tractable two-step procedure—a univariate GARCH is fitted to each asset first, and the pairwise correlations are allowed to evolve afterwards—so that adding assets to the system does not make estimation unmanageable.

Recent studies apply these models to the gold–Bitcoin comparison directly. Snene Manzli et al. (2025), using DCC, ADCC, and GO-GARCH specifications to study gold and Bitcoin against a set of blue-economy and green-finance assets during the COVID-19 and Russia–Ukraine crises, find that gold's time-varying correlation structure is consistent with a diversification and safe-haven role, while Bitcoin's is weaker and more variable. Working with a rolling-window DCC-GARCH on equity, gold, and Bitcoin returns through the COVID-19 and Ukraine episodes, Kumar et al. (2025) reach the same ranking: gold is the more reliable safe haven, whereas Bitcoin shows only weak safe-haven properties and is better described as a diversifier. We follow this literature by estimating a GARCH(1,1) for each asset's volatility and a DCC-GARCH for the time-varying correlations, and we report a 90-day rolling correlation alongside the DCC-GARCH series so that our conclusions about Bitcoin's equity-like co-movement do not depend on a single modelling choice or an arbitrary window length.

2.7 Macro-Financial Linkages and Cointegration

Cointegration analysis has shaped the empirical study of gold and its macro-financial drivers since the work of Johansen (1991). When variables share long-run trends, regression in levels can produce spurious results, and regression in differences can throw away long-run information. Cointegration analysis delivers a middle path: it identifies linear combinations of non-stationary variables that are stationary, and the Vector Error Correction Model uses these combinations to separate the long-run equilibrium from short-run dynamics. Pukthuanthong and Roll (2011) study gold and the major currencies and show that the price of gold is systematically related to currency depreciation in every country, not only to movements in the U.S. dollar.

When several series each wander on their own but remain tied together in the long run, a regression in levels can manufacture a relationship that is not really there—the classic spurious-regression problem—while differencing every series removes the long-run information along with the spurious part. The Johansen test resolves this by estimating how many independent long-run relationships, the cointegrating rank, the system contains, and the resulting VECM splits each variable's movement into a long-run equilibrium term and the short-run dynamics around it. Because our series are integrated of order one, even short-run predictability has to be tested carefully: we use the lag-augmented procedure of Toda and Yamamoto (1995), which keeps a Granger-causality test valid without first differencing the data and throwing away the long-run information.

Recent applied work supports this approach for gold but also warns about its limits. Rana and O'Connor (2023) examine the domestic macroeconomic determinants of precious-metals prices across a range of developed and emerging economies from 1979 to 2020, combining Johansen cointegration tests, the VECM, VAR, and ARDL methods. Where long-run relationships exist, they link precious-metals prices to variables such as inflation, equity prices, interest rates, and the unemployment rate; but their central message is one of heterogeneity. No single technique and no single set of explanatory variables works across all countries and metals, and in some countries no cointegrating relationship is found at all. Two cautions follow for RQ2. First, labour-market and broader macroeconomic conditions—not only inflation and the dollar—belong in a gold-price system, which is consistent with the prominent role we later find for unemployment. Second, any single-country result, including ours, is conditioned on the sample economy and need not transfer unchanged to other monetary regimes.

The lesson from this literature is also one about sampling frequency. Macroeconomic variables are released monthly or quarterly, while gold trades daily. A cointegration system that mixes these frequencies must either downsample to the lower frequency or incur the costs of forward filling, which artificially inflates the effective sample size of the macro variables and compresses their short-run variance. We take the first option and estimate the system at the natural monthly frequency of the macro data.

2.8 Contribution of This Thesis to the Literature

Several of the studies reviewed in this chapter come close to what we do. This section sets out, study by study, where our analysis goes beyond each.

Pflaumer (2024) models gold as a geometric random walk within the ARIMA framework over 1971–2024. Our RQ1 confirms his conclusion on daily data for 2008–

2025 and uses it as a starting point: because gold's own history is uninformative beyond a day, the forecasting question passes to a multivariate system in RQ2. What we add is the link between the two questions, not the replication.

On the determinants side, three strands meet in our RQ2, none of them answering its question on its own. Rana and O'Connor (2023) ask whether long-run relationships exist between precious-metals prices and domestic macro variables across many countries, and find heterogeneity; their tests establish that relationships exist, not how much each variable explains. Robinson (2019) uses the tools we need, impulse responses and variance decomposition on a gold system, but her variables are OECD output, oil, the exchange rate, and the policy rate, so her decomposition says nothing about how inflation ranks against unemployment. Thaver and Lopez (2016) study the pair we care about, gold and U.S. unemployment, but only that pair, so their result shows that unemployment matters without locating it among the other drivers. We assemble the pieces into one system: six variables over a crisis-rich sample, ranked by forecast error variance shares. The answer — unemployment ahead of CPI in gold's twelve-month forecast error, by 33.57 against 7.40 percent in the full sample — is one that none of the three designs could have produced. What we claim is the ordering, not the precise share: it agrees in sign with Thaver and Lopez (2016) and holds up under the robustness checks in Chapter 6, whereas the size of the gap is specific to a window that includes the 2020 pandemic.

Snene Manzli and Jeribi (2024) and Kumar et al. (2025) compare gold and Bitcoin with dynamic-correlation methods and conclude that gold is the more dependable of the two; their samples end in 2022 and at the January 2024 ETF approval, respectively. We reach the same ranking on a sample extended through August 2025, with stress conditioned on VIX quantiles rather than on named crisis windows. The extension does two things at once. Bitcoin's stress correlation with equities (+0.39 in the top quartile, +0.49 in the top decile) confirms their verdict in the era of its deepest institutional integration, while gold's own mildly positive stress correlation (+0.16 at the extreme) qualifies a safe-haven label that the earlier studies leave intact. Wu (2025) documents the structural rise in Bitcoin's equity correlation after its institutional milestones, but gold does not enter his framework. We place both assets inside one, sharing a common sample, GARCH specification, and stress regimes, so that the institutionalization story and the safe-haven comparison come from a single set of estimates rather than from separate papers.

Chapter 7 develops the point these comparisons share: applied to a single seventeen-year dataset, ARIMA, the VECM, GARCH(1,1), and DCC-GARCH give answers that agree, and several findings — the unemployment ranking above all — surface only when the methods are combined.

CHAPTER 3: ECONOMIC UNCERTAINTY AND FINANCIAL STRESS

3.1 Two Related but Distinct Concepts

Economic uncertainty and financial stress are often used interchangeably, but they describe distinct phenomena. Uncertainty refers to the difficulty of predicting the future state of the economy. It increases when policy regimes shift, when confidence in inflation forecasts weakens, or when geopolitical conditions become unstable. Financial stress, by contrast, refers to the strains that become observable in market data once adverse conditions have begun to materialize: borrowing costs rise, interbank lending tightens, equity-market volatility increases, and investors reduce their exposure to risk.

The two also differ in how they are measured. Economic uncertainty is hard to observe directly; it is usually proxied by survey dispersion, forecaster disagreement, or news-based indices that track uncertainty-related language in the press. Financial stress is priced in markets in close to real time — it shows up in the VIX, in credit spreads, and in the cost of interbank funding — so it can be read off observable series day by day. Sokhanvar and Hammoudeh (2024) show the practical consequence: a shock to the VIX moves asset prices immediately and then fades over the following days, so a stress measure is most informative when used at the moment stress is elevated rather than averaged across years. We therefore use the VIX to mark out high-stress days in RQ3, while the slower-moving macro variables in RQ2 stand in for the broader uncertainty that builds up over months.

The two concepts are mutually reinforcing. Uncertainty typically precedes stress, as investors demand higher compensation for risk and reduce their holdings of risky assets. Once stress emerges, it raises the probability of further adverse outcomes and so amplifies the original uncertainty. Gold tends to attract demand in both environments. The reason is structural: gold's value does not depend on any counterparty's ability to meet an obligation, and its supply cannot be expanded at the discretion of a central bank.

The distinction lines up with the two empirical tests that follow. Slow-moving economic distress — weakening growth, rising unemployment, deteriorating confidence — is the uncertainty that the monthly macro-financial analysis of RQ2 is built to capture. Acute, market-priced stress — the sudden repricing of risk during a crash — is what the volatility and correlation analysis of RQ3 captures through the VIX. The two are linked: recent connectedness studies of the COVID-19, Russia–Ukraine, and Silicon Valley Bank episodes show how a shock in one market is transmitted rapidly across assets and borders (Snene Manzli and Jeribi, 2024), so an episode that begins as localized stress can widen into broad uncertainty within days. Separating them shows which kind of protection gold actually provides, which RQ2 and RQ3 answer in turn.

3.2 Hedge, Safe Haven, and Speculative Asset

Three terms recur throughout this thesis and require precise definition. We follow Baur and Lucey (2010).

A hedge is an asset that offsets movements in another asset. If an investor holds a portfolio of equities that declines by 10 percent, a hedge is an asset that does not decline with it, or that appreciates. On average, a hedge is either uncorrelated with or negatively correlated to the asset it hedges, measured over the whole sample rather than in any single episode. Gold has historically served as a hedge against equities in this average sense, and the near-zero full-sample correlation between gold and the S&P 500 that we report later is consistent with that role. A weaker category, the diversifier, sits between a hedge and no protection at all: it is positively but imperfectly correlated with the asset on average, so it reduces portfolio risk in normal times but offers no guarantee during a crisis.

A safe haven is a stronger condition than a hedge. It must preserve its value specifically during periods of market crisis, not simply on average. An asset may function as an effective hedge in calm periods but fail to protect investors during a severe downturn, and the converse is also possible. This is why safe-haven status has to be tested with stress-conditional or time-varying methods rather than a single full-sample correlation, as the recent literature surveyed in Chapter 2 insists (Snene Manzli and Jeribi, 2024; Sokhanvar and Hammoudeh, 2024). Gold has historically served as both a hedge and a safe haven, separating it from many other assets—a combination we test rather than assume.

A speculative asset serves the opposite function. Investors hold it in the expectation of price appreciation rather than for protection against losses. Speculative assets are generally characterized by large price fluctuations, positive co-movement with equities, and an increase in portfolio risk when added to a diversified portfolio. The recent evidence on Bitcoin fits this description: Wu (2025) shows that Bitcoin's correlation with equities has risen with institutional adoption, and Acikgoz (2025) finds its behaviour dominated by volatility rather than by any reliable defensive property.

In Chapter 6, we test which of these three categories best describes Bitcoin. The distinction has practical consequences. Investors who buy gold generally seek protection, whereas investors who buy equities generally seek returns, and the two motives call for different assets. An investor who acquires Bitcoin believing it provides gold-like protection, when the data indicate it behaves like a risk asset, holds an instrument mismatched to its intended purpose, and the mismatch becomes visible only during a downturn, when protection is needed most.

3.3 Three Stress Episodes in the Sample

The 2008–2025 window contains three stress episodes of different origin, which together test the hedge, safe-haven, and speculative-asset framework against very different shocks. Each strained the financial system through a different channel — banking and credit in 2008, public health in 2020, and inflation in 2022 — so an asset that protected investors across all three has a stronger claim to safe-haven status than one that performed well in only one. The episodes also span the entire traded history of Bitcoin, setting gold's long-established behaviour against Bitcoin's still-evolving profile. Recent connectedness studies covering these crises confirm that they were genuine system-wide stress events rather than isolated shocks (Snene Manzli and Jeribi, 2024).

3.3.1 Global Financial Crisis, 2008–2009

Lehman Brothers failed in September 2008. Credit markets froze, the S&P 500 declined by more than 30 percent, and the Federal Reserve aggressively reduced policy rates. Gold rose from approximately $700 per ounce at its lows to above $1,800 by 2011. The crisis also initiated an extended period of unconventional monetary policy that influenced the remainder of the sample. As a test of gold's safe-haven role against a credit-driven crisis, this is the clearest episode in the early part of the sample, and gold came through it. Bitcoin did not yet exist as a traded asset, so this episode is a test of gold alone. The pattern it shows—rising demand for a default-free store of value while the banking system was under strain—recurs in the later episodes.

3.3.2 COVID-19, 2020

The pandemic produced the fastest equity-market decline in modern financial history, followed by a comparably rapid recovery. The unemployment rate reached 14.8 percent in April 2020 and returned below 4 percent within two years. Gold exceeded $2,000 per ounce for the first time in August 2020. Bitcoin declined alongside equities in March 2020 and subsequently rose to a speculative peak in late 2021. This episode provides the clearest test of Bitcoin's safe-haven claim and is the principal motivation for RQ3. The VIX reached its sample maximum of about 83 during the March panic. The two assets parted ways: after a brief dash-for-cash dip, gold regained its footing and reached a record high by August, while Bitcoin tracked the equity sell-off down before later rallying as a risk asset.

3.3.3 Inflation Shock, 2022–2023

U.S. CPI inflation peaked above 9 percent in mid-2022. The Federal Reserve raised the federal funds rate from near zero to above 5 percent in the most rapid tightening cycle since the early 1980s, and the 10-year Treasury yield rose from below 2 percent to above 4.5 percent. Gold retained its value through the tightening cycle and resumed rising in 2024–2025, surpassing $3,400 per ounce. This resilience is consistent with the regime-dependent inflation-hedging behaviour documented by Valadkhani and O'Mahony (2024), in which gold's protective role is strongest in high-inflation regimes. Bitcoin, by contrast, declined with equities in 2022, a pattern more consistent with speculative behavior than with hedging. Gold's ability to hold its value through the most aggressive tightening in four decades—when rising real yields would normally weigh on a non-yielding asset—points to demand driven by broad uncertainty rather than by inflation alone, and it previews the RQ2 finding that gold responds more to economic distress than to consumer prices in isolation.

CHAPTER 4: THE DATASET

4.1 Sources and Coverage

Our raw sample runs from January 2, 2008, to August 29, 2025. The data are drawn from two public, authoritative sources: the financial-market price series from Yahoo Finance, and the macroeconomic series from the Federal Reserve Economic Data (FRED) service maintained by the Federal Reserve Bank of St. Louis.[1] The dataset contains the gold price, the U.S. Dollar Index, crude oil, silver, the S&P 500, Bitcoin, the federal funds rate, the 10-year Treasury yield, CPI, and the unemployment rate. The financial-market series (gold, silver, crude oil, the S&P 500, Bitcoin, the USD Index) are daily, while the macroeconomic series (CPI, the federal funds rate, the 10-year Treasury yield, and the unemployment rate) are monthly; price series are end-of-day values in U.S. dollars.

We also include the official CBOE Volatility Index (the VIX) from FRED, series VIXCLS, matched to the other series by date. Every series is published by an official source, so the dataset can be reconstructed in full from FRED, Yahoo Finance, and the relevant exchanges.

Figure 4.1 shows the seventeen-year gold price series that anchors the entire thesis.

Figure 4.1: Daily spot gold price, January 2008 to August 2025. The red dashed line marks the train/test split for the ARIMA exercise in Chapter 6.

[1]https://fred.stlouisfed.org. Series retrieved in September 2025.

4.2 Variables

Table 4.1 lists the eleven variables, their units, and their frequency.

Table 4.1: Variables in the dataset.

Variable	Description	Unit	Freq.
Gold_XAUUSD	Spot gold price per troy ounce	USD/oz	Daily
USD_Index_DXY	U.S. Dollar Index	Index	Daily
Crude_Oil_WTI	WTI crude oil price	USD/bbl	Daily
Silver_XAGUSD	Spot silver price	USD/oz	Daily
SP500	S&P 500 stock market index	Points	Daily
Bitcoin_BTC-USD	Bitcoin price in USD (from 9/17/2014)	USD	Daily
Interest_Rate	Federal Funds Rate	%	Monthly
10Y_Treasury_Yield	10-year Treasury note yield	%	Monthly
Inflation_CPI	Consumer Price Index	Points	Monthly
Unemployment_Rate	U.S. unemployment rate	%	Monthly
VIX_CBOE	CBOE VIX index (from FRED)	Index	Daily

Three of these series — silver, crude oil, and the federal funds rate — are reported for context and descriptive completeness; the formal models use gold, the USD Index, the S&P 500, CPI, the 10-year Treasury yield, the unemployment rate, and the VIX.

4.2.1 Why Each Variable Is in the Model

Gold is our main variable of interest. Supply grows slowly through mining, and there is no issuer that can default, which is why gold tends to hold its value when other stores of wealth do not.

The USD Index (DXY) measures the dollar against six major currencies. Because gold trades in dollars, a stronger dollar makes gold more expensive for foreign buyers and usually pushes the dollar price down.

The S&P 500 is a proxy for how U.S. stock investors feel about risk. A negative or near-zero correlation with gold fits the hedge case. A strong positive correlation does not.

Bitcoin enters our dataset in September 2014. Commentators often call it digital gold. Whether the data support that label is the subject of RQ3.

The federal funds rate and the 10-year Treasury yield measure the opportunity cost of holding gold, which pays no income. Higher rates make interest-bearing assets more attractive and tend to push gold down. Lower rates have the opposite effect.

CPI tracks the erosion of purchasing power. The standard view is that gold holds its real value during periods of high inflation because we cannot expand the gold supply the way central banks can expand currency supply.

The unemployment rate is a lagging indicator of the business cycle. When unemployment rises, the economy is weakening, and money tends to flow into defensive assets.

The VIX is the CBOE's measure of expected stock-market volatility over the next 30 days, often called the market's "fear gauge." It rises when investors expect trouble ahead and falls when things are calm. Including it alongside the other variables gives us a direct read on market fear, rather than the realized economic conditions captured by unemployment and CPI.

4.3 Descriptive Statistics

Table 4.2 gives descriptive statistics for the full sample. The modeling in Chapter 6 has to account for these numbers.

Table 4.2: Descriptive statistics, full sample.

Variable	Mean	Median	Std Dev	Min	Max
Gold (USD/oz)	1,534.19	1,364.10	498.70	704.90	3,473.70
USD_Index	91.03	93.51	9.94	71.30	114.05
Oil (USD/bbl)	72.82	72.15	22.21	−37.63	145.29
Silver (USD/oz)	21.65	19.50	6.83	8.79	48.58
SP500 (points)	2,662.58	2,180.39	1,436.35	676.53	6,501.86
Bitcoin (USD)	24,540.11	10,214.38	29,111.84	178.10	123,344.06
Interest_Rate (%)	1.324	0.200	1.729	0.050	5.330
10Y_Treasury_Yield (%)	2.665	2.600	0.974	0.620	4.770
Inflation_CPI (pts)	252.19	241.74	31.97	211.40	322.13
Unemployment (%)	5.935	5.100	2.244	3.400	14.800
VIX_CBOE (index)	19.996	17.550	9.024	9.140	82.690

Several figures are worth noting. Gold's mean is \$1,534 against a maximum of \$3,474, so most of the appreciation happened late in the sample. Bitcoin's standard deviation (\$29,112) is larger than its mean (\$24,540), the mark of a highly volatile asset. Oil's minimum of −\$37.63 is the single-day collapse in April 2020, which any correlation analysis has to treat carefully. The VIX averages 20 across the sample but peaks at 82.69, at the height of the March 2020 panic. CPI rises steadily across the whole sample, while unemployment rises and falls with the business cycle.

We plot all six VECM variables on a common standardized scale (Figure 4.2) to make the broader patterns visible: the steady upward drift in gold, the S&P 500, and CPI; the business-cycle behavior of unemployment, and the marked dispersion that opens up after 2020.

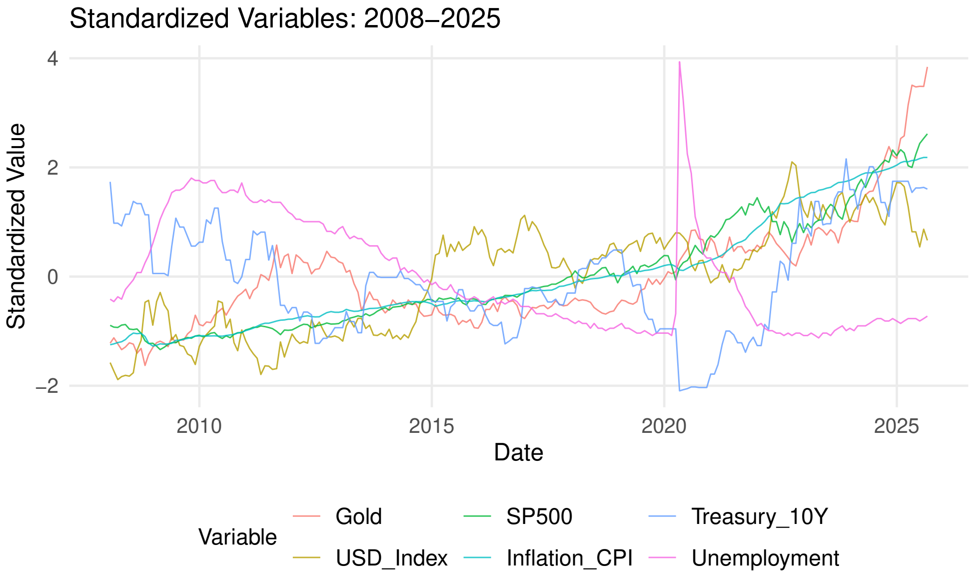

Figure 4.2: Standardized levels of the six VECM variables, 2008–2025. Each series is rescaled to mean zero and unit variance to make the trends visually comparable.

4.4 Possible Correlations and Their Interpretation

The raw level correlations give us a first impression of the structure of the dataset:

Table 4.3: Selected level correlations, full sample.

Pair	Correlation	Interpretation
Gold vs Inflation_CPI	0.84	Gold rises with CPI over the long run
Gold vs SP500	0.83	Both trend upward through the expansion
Gold vs USD_Index	0.49	Positive in levels, driven by 2022–2025
Gold vs Treasury_10Y	0.22	Weak positive in levels
Gold vs Unemployment	−0.28	Negative with labor-market weakness
SP500 vs Inflation_CPI	0.97	Both trend upward strongly

The full pairwise picture is in Figure 4.3, where unemployment is the only variable that lines up negatively with the others.

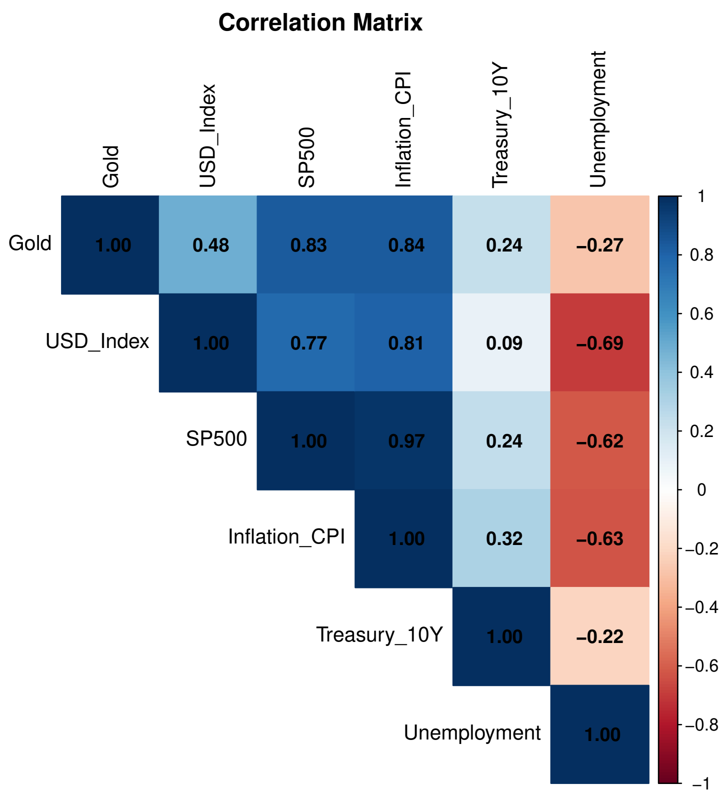

Figure 4.3: Correlation matrix of the six VECM variables in levels. High correlations among gold, the S&P 500, and CPI reflect a common upward trend across all three series in the sample.

These correlations are large — gold and CPI at 0.84, gold and the S&P 500 at 0.83 — but they are not evidence of economic causality or hedging behaviour. These are correlations of variables in levels, and all three series (gold, CPI, S&P 500) have been rising steadily for most of the past seventeen years. Any two upward-trending series will show a high correlation, whether or not they are related. The table is therefore descriptive only: it summarizes co-movement in the raw data and is used nowhere in the thesis as evidence of a relationship. For judging how assets actually move together — the question behind a hedge or safe haven — correlations of returns are far more informative than correlations of levels, and those are what we use in RQ3. The level relationships that matter for RQ2 are instead isolated with the VECM, which separates genuine long-run links from shared trends.

4.5 Data Preparation

A few points about how we prepared the data. First, Bitcoin prices began on September 17, 2014, so any analysis that includes Bitcoin uses the shorter 2014–2025 window rather than the full 17 years. Second, four macro variables — CPI, unemployment, the federal funds rate, and the 10-year Treasury yield — are available only once a month. The financial variables trade every day. For descriptive purposes, we aligned the monthly values with daily observations by carrying the last available release forward, which is the standard convention. For the VECM in RQ2, however, we estimate the system at monthly frequency by sampling the daily market variables on the last trading day of each month. This is the natural release frequency of the macro variables and avoids the artificial inflation of effective sample size that forward-filling produces in daily data. We discuss the choice in more detail in Chapter 5.

Third, the assembled file contained a small number of placeholder rows dated to the first calendar day of a month that fell on a weekend or market holiday, on which the daily market prices were simply carried forward from the previous trading day. Because these rows are not genuine trading observations — gold and the S&P 500 register an exactly zero return on them while Bitcoin, which trades continuously, does not — we remove them so that the daily panel contains trading days only. Bitcoin returns on the following trading day accordingly span the intervening non-trading days, the standard convention when a continuously traded asset is aligned to an exchange calendar. After this step the daily sample used for RQ3 contains 2,753 trading days from the start of Bitcoin coverage through August 2025, and the monthly RQ2 panel is unaffected at 212 observations, since end-of-month sampling never selected the removed rows.

For the RQ3 return, volatility, and correlation analysis, the series are aligned to a common set of exchange trading days: weekend and market-holiday rows are excluded, so gold, the S&P 500, and Bitcoin are compared on the same days rather than each on its own calendar. Bitcoin trades continuously, but its return on the first trading day after a weekend or holiday is measured over the gap, so no Bitcoin observation is discarded; it is simply dated to the next common trading day. No interpolation or back-filling of missing prices is used anywhere in the analysis: the only adjustment is the removal of the synthetic carried-forward rows described above, after which every remaining row is a genuine trading-day observation for all three assets. This matters because spurious zero-return days would otherwise attenuate the rolling correlations and the DCC-GARCH estimates and distort the volatility comparison.

CHAPTER 5: MODELS USED

5.1 Matching Questions to Models

Each research question needs a different tool. RQ1 asks whether past gold prices predict future gold prices, which is a one-variable problem that ARIMA handles well. RQ2 asks how several macro and financial variables move together with gold. Those variables are nonstationary and exhibit long-run relationships, which is exactly the setup the Vector Error Correction Model (VECM) was designed to handle. RQ3 is about volatility and correlation, both of which change over time in financial data. GARCH(1,1) handles volatility. DCC-GARCH extends this to allow correlations to change over time. Table 5.1 summarizes which model does which job.

Table 5.1: Mapping of research questions to models.

Research question	**Empirical task**	**Model**
RQ1	Univariate forecasting	ARIMA
RQ2	Long-run macro-financial linkages	VAR / VECM
RQ3	Time-varying volatility	GARCH(1,1)
RQ3	Time-varying correlation	DCC-GARCH

5.2 ARIMA for RQ1

ARIMA stands for Autoregressive Integrated Moving Average. It is the standard model for forecasting a single series using only its own past values. The (p, d, q) after the name refers to three choices we have to make: how many past values to include (p), how many times to difference the series to make it stable (d), and how many past forecast errors to include (q). In equation form,

$$(1 - \phi_1 L - \cdots - \phi_p L^p)\,(1 - L)^d\, y_t = (1 + \theta_1 L + \cdots + \theta_q L^q)\,\varepsilon_t. \tag{5.1}$$

Before we fit an ARIMA, the series has to be stationary, meaning its statistical behavior does not drift over time. Gold is not stationary in levels: the price keeps climbing, so we have to difference it first. Differencing means working with the change from one day to the next instead of the price itself. We call a series I(1) when it becomes stationary after one round of differencing. The ACF and PACF plots of the differenced series help us pick the orders p and q, and information criteria like AIC or BIC let us choose among candidate models. After we fit a model, the residuals should look like random noise with no pattern left. We check this with the Ljung-Box test and the residual ACF.

An unrestricted automatic AICc search using auto.arima() prefers ARIMA(0,1,4) with drift. Its AICc improvement over the simpler ARIMA(0,1,0) is small, however — under five points over more than three thousand observations — and the additional moving-average coefficients are economically tiny. The same is true within the narrower candidate grid we also fit by AICc, where the lowest-scoring alternative, ARIMA(1,1,0), beats the random walk by only about one AICc point. We therefore retain ARIMA(0,1,0) as the central benchmark because it represents the random-walk hypothesis directly and gives us a transparent reference point. Reporting ARIMA(0,1,0) is also the more conservative choice because it does not overstate the predictive content of gold's own past values.

ARIMA fits RQ1 because daily gold prices carry strong momentum day to day. The model uses only the price history and gives us point forecasts with proper confidence bands. Its limitation is that it ignores everything outside the gold price itself, which is exactly why we need RQ2.

5.3 VAR and VECM for RQ2

A Vector Autoregression (VAR) is ARIMA for several series at once. Each variable depends on lags of itself and lags of every other variable in the system,

$$Y_t = A_1 Y_{t-1} + \cdots + A_p Y_{t-p} + \varepsilon_t, \tag{5.2}$$

where Y_t is the group of variables at time t. A VAR only works when every series in it is stationary. If the series are not stationary, we still have one useful possibility: even though each individual series drifts on its own, they may share a long-run equilibrium that ties them together. We call this cointegration. The idea is that two variables can wander over time, but stay within a certain distance of each other in the long run. The Johansen procedure is the formal test for cointegration and reports a number called the rank, labeled r, which counts how many such long-run relationships exist.

When $r > 0$, the right model is a Vector Error Correction Model (VECM):

$$\Delta Y_t = \alpha\, \beta' \, Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \cdots + \varepsilon_t. \tag{5.3}$$

The term $\beta' Y_{t-1}$ is the long-run equilibrium that the variables share. The term α measures how quickly each variable pulls back toward that equilibrium when it drifts away. A fitted VECM produces three kinds of output. Impulse response functions (IRFs) ask: if one variable gets shocked, what happens to every other variable over the next several periods? Forecast error variance decomposition (FEVD) breaks down the total forecast uncertainty for a variable at a given horizon into shares contributed by each shock source in the system, ranking the drivers by how much they actually explain rather than just whether they move the variable in the expected direction. Granger causality tests check whether one variable helps predict another in the short run.

We estimate the VECM at monthly frequency rather than daily frequency. The macroeconomic variables in the system (CPI, unemployment, and the 10-year Treasury yield) are naturally monthly series. In a daily dataset, these series stay constant within each calendar month and only jump at month boundaries. A VECM that treats each daily row as an independent observation therefore artificially inflates the effective sample size of the macro variables and compresses their short-run variance. We remove this distortion by sampling the daily market variables (gold, the U.S. Dollar Index, the S&P 500) at the last trading day of each month, so that all six variables sit on a common monthly grid. This is the natural release frequency of the macro variables and is the appropriate frequency for studying their relation with gold.

5.4 GARCH(1,1) for RQ3

One well-known pattern in financial data is that wild days tend to cluster. After a big move in the market, the next few days also tend to see big moves, whether up or down. In other words, volatility itself changes over time. A standard model like ARIMA assumes volatility is constant, which is wrong for daily returns. The GARCH model addresses this. GARCH stands for Generalized Autoregressive Conditional Heteroskedasticity: it allows the conditional volatility of a series to depend on past shocks and past volatility. The GARCH(1,1) version lets today's variance depend on yesterday's squared surprise and yesterday's variance,

$$\sigma_t^2 = \omega + \alpha\,\varepsilon_{t-1}^2 + \beta\,\sigma_{t-1}^2. \tag{5.4}$$

The conditions $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$, and $\alpha + \beta < 1$ keep the variance well-behaved over time. We model the residuals with a Student-t distribution to allow for the fat tails that show up in daily returns, meaning extreme moves happen more often than a normal bell curve would predict. We use that specification throughout the thesis.

GARCH(1,1) yields a single volatility time series for each asset, allowing us to compare gold, the S&P 500, and Bitcoin on the same scale.

5.5 DCC-GARCH for RQ3

Engle (2002)'s Dynamic Conditional Correlation (DCC) model extends GARCH to multiple assets. First, we fit a GARCH to each asset separately to get its volatility over time. Then the model lets the correlation between each pair of assets also change over time, driven by how each one has been moving lately. The two-step setup keeps the math manageable even when we work with several assets.

The output we care about is the time series of correlations between each pair. If a pair's correlation stays near zero or turns negative during stressful market periods, that fits the safe-haven pattern. If the correlation goes up during stress, it does not. We prefer DCC-GARCH to simple rolling correlation because it jointly models volatility and correlation, and it does not force us to pick an arbitrary window length, such as 30 or 90 days. We report both 90-day rolling correlations and DCC-GARCH correlations as a double-check. They give us the same answer. The DCC-GARCH estimates are nonetheless model-dependent: they can be sensitive to the assumed error distribution, the length of the sample, and extreme observations, so we treat the correlation series as indicative of broad patterns rather than as precise point estimates, and we lean on the agreement between the DCC-GARCH and the model-free rolling correlation rather than on either alone.

5.6 Diagnostic Logic

Our empirical strategy includes diagnostic checks because financial data often violates the assumptions of ideal models. We check ARIMA residuals for remaining autocorrelation. We check VECM residuals for serial correlation, normality, ARCH effects, and system steadiness using characteristic roots. We fit GARCH models with a Student-t distribution because daily financial returns have fat tails. These diagnostics do not remove all limitations, but they make our interpretation more transparent. Where multivariate normality fails — which occurs frequently in long financial samples — we lean on parameter signs and economic plausibility rather than on the exact distribution of the residuals.

CHAPTER 6: EMPIRICAL FINDINGS

In this chapter, we run each model on the data and report the results. Every number below comes from the R code on the dataset from Chapter 4. We show model choices, diagnostic checks, and performance measures alongside each result.

6.1 RQ1: Forecasting Gold from Its Own Price History

A long-standing question in finance is whether the past price of an asset tells us anything useful about its future price. If it does, the price history alone is enough to forecast; if it does not, the current price already reflects what is known, and forecasting has to bring in other variables. RQ1 asks this question of gold over the period 2008 to 2025.

Our tool is the ARIMA model, which forecasts a series from its own past values and past forecast errors. ARIMA requires the series to be stationary, meaning its mean and variance do not drift over time. Gold is not stationary in levels: the price starts near \$700 and ends above \$3,400. The augmented Dickey–Fuller (ADF) test confirms this, failing to reject a unit root in levels ($p > 0.99$). After first differencing, the series becomes stationary (ADF $p < 0.01$), so gold prices are integrated of order one, I(1).

Figure 6.1 shows the same thing visually. In levels, the ACF decays slowly, which is the signature of a non-stationary series. After differencing, the autocorrelations fall inside the confidence bands and no autoregressive or moving-average structure is left.

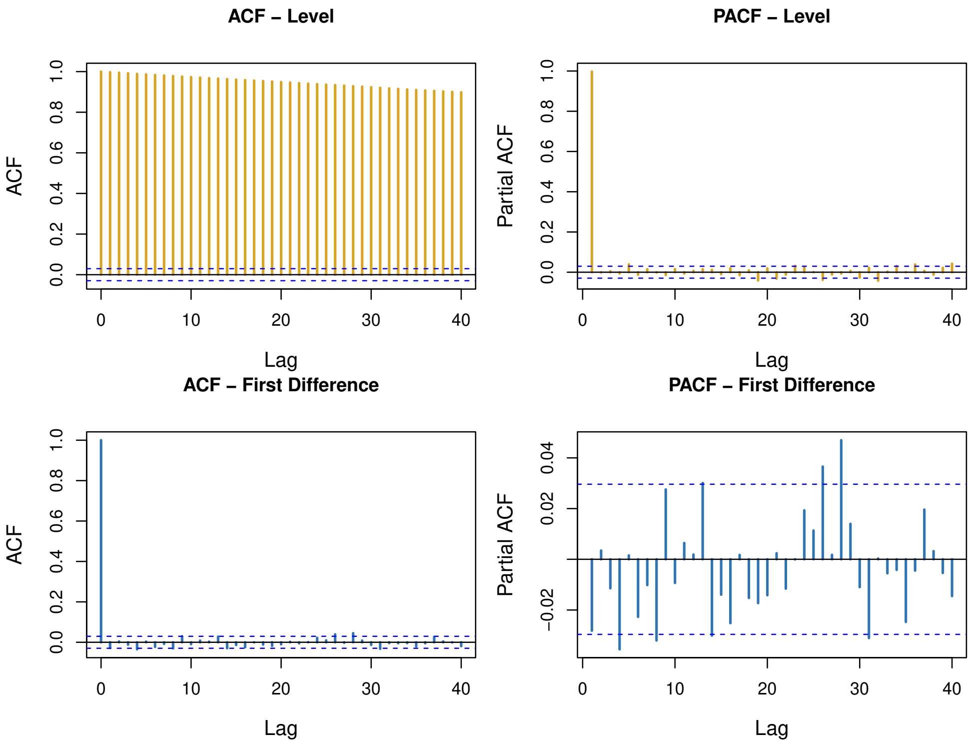

Figure 6.1: ACF and PACF of gold prices in levels (top) and first differences (bottom). The slow decay in the level ACF confirms non-stationarity; the first-differenced series shows no remaining linear structure, which is consistent with a random walk.

We then choose the model order. An unrestricted auto.arima() search prefers ARIMA(0,1,4) with drift, and within a narrower candidate grid ranked by AICc the lowest-scoring specification is ARIMA(1,1,0); but both improve on the simpler ARIMA(0,1,0) by only a marginal amount — under five AICc points for the wider search and about one point for ARIMA(1,1,0) — and the extra terms are economically negligible. We therefore keep ARIMA(0,1,0) as our benchmark. This specification has no autoregressive or moving-average terms and one order of differencing, which makes it a pure random walk: the best forecast of tomorrow's gold price is simply today's price plus noise. Read this way, the price history — including the 2020 rally and the 2022 inflation shock — adds nothing to the forecast of the next observation. That is what we would expect in a deep and liquid market, where any exploitable pattern is traded away before it shows up in the data.

Before relying on the model, we check that its residuals behave as the specification assumes. The diagnostics in Figure 6.2 support it: the residual ACF stays within the confidence bands, the histogram is roughly centred on zero, and the only real departure from the model's assumptions is a tail heavier than the normal benchmark. That is typical of daily returns, and we deal with it through the Student-t GARCH in RQ3.

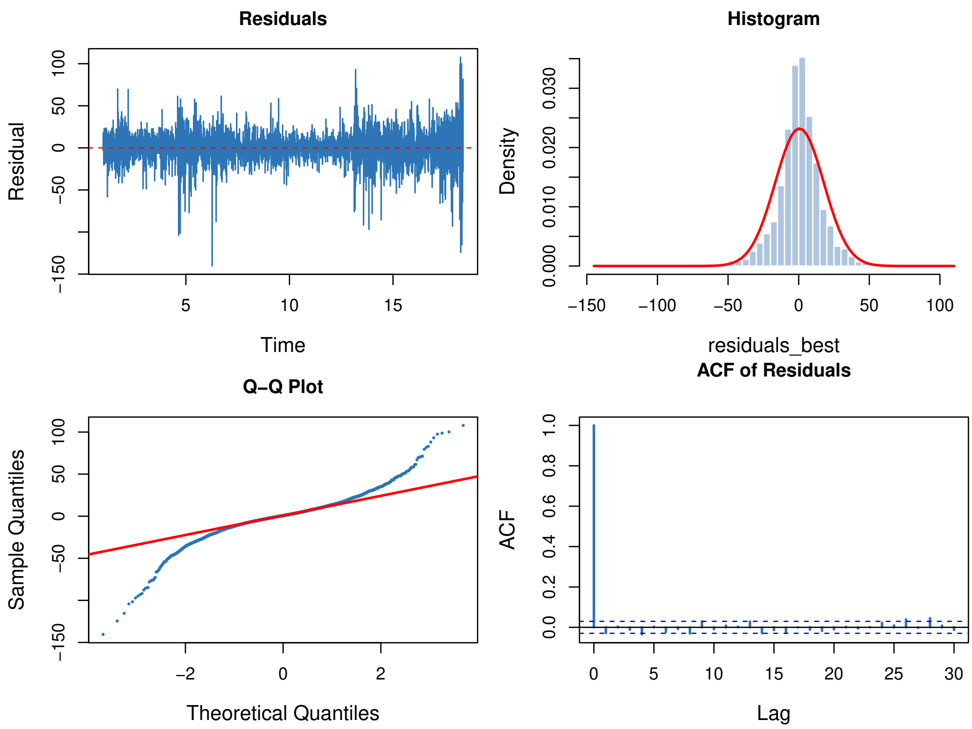

Figure 6.2: Residuals from ARIMA(0,1,0). The residuals fluctuate around zero with no autoregressive structure visible in the ACF, but heavier tails than the normal benchmark, especially around the 2020 stress episode.

With the random-walk specification settled and its fit confirmed, we test how well it forecasts. We do this in two ways: a fixed-origin 60-day forecast, which projects a single path 60 days ahead, and a rolling one-step-ahead forecast, which re-estimates one day at a time. Table 6.1 reports both, using the root mean squared error (RMSE), the mean absolute error (MAE), and the mean absolute percentage error (MAPE). The first two are measured in dollars and have to be read against gold's price level of roughly \$3,300; MAPE restates the error as a percentage of the actual price,

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \times 100\%, \tag{6.1}$$

where A_t is the actual price and F_t the forecast at time t. For each day it takes the gap between the forecast and the actual price, divides it by the actual price to turn it into a percentage, and then averages these percentages over all the days in the forecast. The result is the average size of the forecast error expressed as a percentage of price, so a MAPE of 0.94 percent means the forecast was off by about one percent of the gold price on average. Because it is a percentage rather than a dollar figure, it can be read on its own without reference to the price level.

The rolling one-step-ahead forecast has a MAPE of 0.67 percent, about as low as a univariate model of daily gold can reach, and it earns this by using the most recent

Table 6.1: ARIMA(0,1,0) forecast accuracy.

Convention	RMSE	MAE	MAPE
Fixed-origin (60-day)	\$40.38	\$31.81	0.94%
Rolling 1-step-ahead	\$29.19	\$22.44	0.67%

observed price at each step. The 60-day fixed-origin forecast looks almost as good at 0.94 percent, but this number is misleading. Because the fitted model is a random walk, its 60-day forecast is a single flat value equal to the last observed price of \$3,350. Over those 60 days the actual price only drifted between \$3,275 and \$3,475 around that line, so the small error reflects a quiet market rather than a model that forecast anything. A low MAPE here measures how little the price moved, not how well the model predicted it.

Figure 6.3 makes the point directly: the forecast is a horizontal line at \$3,350 with widening uncertainty bands, while the realized price drifts around it.

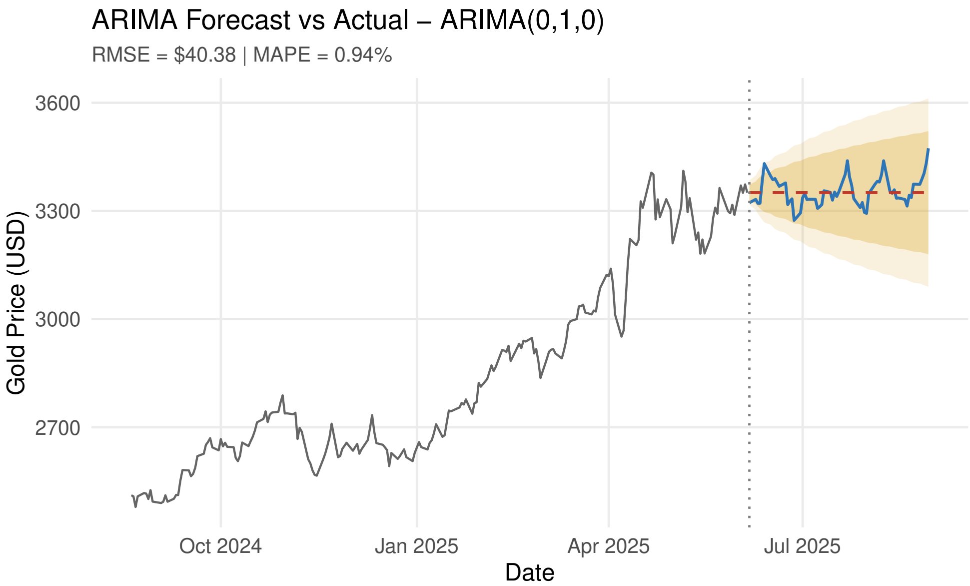

Figure 6.3: ARIMA(0,1,0) 60-day forecast against the realized gold price. The forecast is a flat line at the last observed value (\$3,350) while the actual price wandered between \$3,275 and \$3,475.

The rolling one-step-ahead forecast in Figure 6.4 shows the opposite case. Because each prediction is anchored to the most recent observed price, the forecast follows the realized series closely, and the rolling RMSE falls to \$29.19 with a MAPE of 0.67 percent.

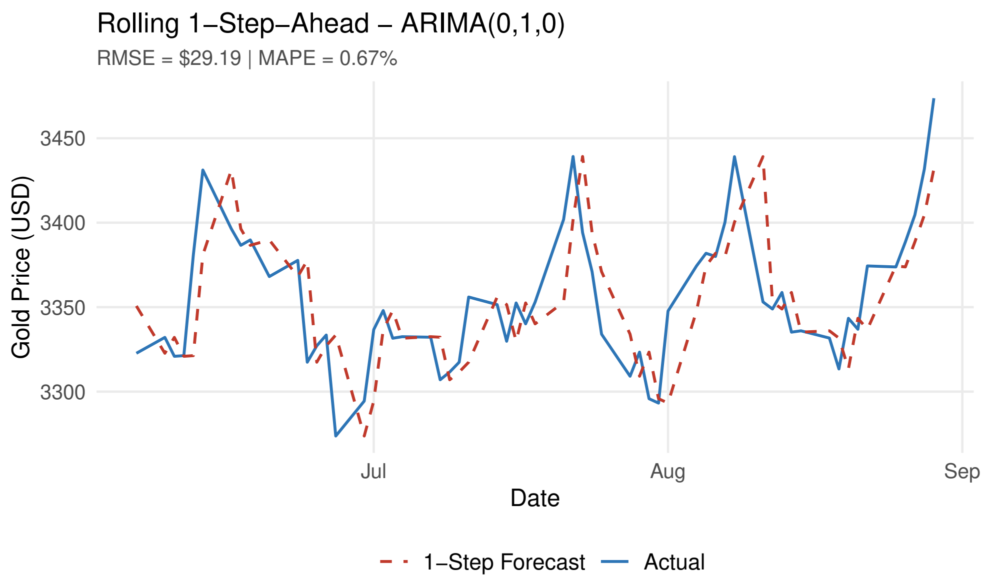

Figure 6.4: Rolling one-step-ahead ARIMA(0,1,0) forecast versus realized gold price. RMSE \$29.19, MAPE 0.67%. The forecast tracks the actual series tightly because each prediction uses the most recent observed price.

The answer to RQ1 therefore has two parts. At a one-day horizon, the most recent price is a sufficient forecast. At a 60-day horizon, the model still applies but cannot tell one month from another, and its flat forecast shows the limit of using gold's own history rather than anything about gold itself. The precise claim is narrow: gold prices are close to a random walk when forecast from their own past values alone. This does not mean gold is unpredictable under every model — the rest of the thesis shows that other variables carry information about gold over longer horizons — only that gold's own price history is uninformative beyond a day, which is what motivates bringing in macroeconomic variables in RQ2.

6.2 RQ2: Macro-Financial Drivers of Gold

RQ1 left us with a clear gap. Gold's own price history tells us nothing useful beyond a single day, so if we want to understand what actually moves gold over longer horizons, we have to look outside the gold price itself. That is the task of RQ2: to identify which macroeconomic and financial forces drive gold, and to measure how much each one contributes.

We consider five candidate drivers, each with a clear economic reason for inclusion: the U.S. dollar (gold is priced in dollars), the S&P 500 (gold is often held as an alternative to equities), inflation (the classic hedging motive), the 10-year Treasury yield (the income an investor gives up by holding gold, which pays none), and the unemployment rate (a broad gauge of economic distress). We keep the system to gold plus these five so that it stays parsimonious enough for our 212-month sample. We leave out the federal funds rate, whose information is already largely contained in the closely related 10-year yield, and the VIX, which we reserve for RQ3 to define stress regimes rather than use as a driver here.

The natural tool for this question is the Vector Error Correction Model (VECM), which we introduced in Chapter 5. Its usefulness here is specific. All six of our series drift upward over time, and as we saw in Chapter 4, any two upward-trending series look correlated whether or not they are genuinely related. A VECM avoids this trap by separating the stable long-run relationship that ties the variables together from the short-run, month-to-month movements around it, so the connections it reports are not just an artefact of shared trends. We estimate it at monthly frequency, because three of the drivers — CPI, unemployment, and the 10-year yield — are only published monthly; rather than invent daily values for them, we sample the daily market variables (gold, the dollar index, and the S&P 500) on the last trading day of each month. This gives a clean panel of 212 monthly observations from January 2008 to August 2025.

6.2.1 Stationarity and Lag Selection at Monthly Frequency

Before estimating the model, we confirm that it is the right one for the data. A VECM is appropriate only under two conditions: the series must be non-stationary (otherwise a simpler model would do), and they must be cointegrated, meaning that although each wanders on its own, they share one or more long-run relationships that hold them together. We test each in turn.

The first condition holds. Table 6.2 reports the augmented Dickey–Fuller test for each variable: in levels, none can reject the presence of a unit root, but after first differencing, all of them become stationary. Every series is therefore integrated of order one.

Table 6.2: ADF tests on monthly levels and first differences. The tseries implementation truncates reported p-values at 0.01 and 0.99, so these are interpreted as bounds rather than exact values.

Variable	ADF level p	Stationary	ADF diff p	Stationary
Gold	> 0.99	No	< 0.01	Yes
USD_Index	0.0819	No	< 0.01	Yes
SP500	> 0.91	No	< 0.01	Yes
Inflation_CPI	> 0.90	No	< 0.01	Yes
Treasury_10Y	0.6864	No	< 0.01	Yes
Unemployment	0.0565	No	< 0.01	Yes

The second condition holds as well. The AIC points to two lags for the monthly system, which is sensible for monthly macroeconomic data. The Johansen test — the standard test for how many long-run relationships a set of series shares — is reported in Table 6.3. Both its trace and maximum-eigenvalue versions find at least two such relationships at the 5 percent level. We therefore estimate the system as a VECM with cointegrating rank $r = 2$.

Table 6.3: Johansen cointegration tests at monthly frequency.

Test	$r \leq 0$	$r \leq 1$	$r \leq 2$	$r \leq 3$	$r \leq 4$	$r \leq 5$
Trace statistic	145.23	76.17	38.85	22.17	11.23	3.85
5% critical value	102.14	76.07	53.12	34.91	19.96	9.24
Max-eigenvalue statistic	69.07	37.32	16.68	10.94	7.39	3.85
5% critical value	40.30	34.40	28.14	22.00	15.67	9.24

The trace statistic at $r \leq 1$ (76.17) exceeds its 5 percent critical value (76.07) by only a narrow margin, so we also examined the conservative rank $r = 1$ as a sensitivity check. Under $r = 1$, unemployment remains the largest macro contributor to gold's twelve-month forecast error variance (10.7 percent), marginally ahead of CPI (10.1 percent); the qualitative ranking in which unemployment has the largest macro contribution is preserved, although the magnitude of this contribution — about a third of the forecast error — is specific to the $r = 2$ specification selected by the Johansen tests.

6.2.2 Gold Equation in the Monthly VECM

With the model in place, we first look at the equation for gold itself. Table 6.4 reports its coefficients. The error-correction term ect2 is significant, which tells us gold does adjust back toward its long-run relationship with the other variables after it drifts away, confirming the system is genuinely tied together rather than moving at random. Among the short-run terms, lagged unemployment is positive and significant at the 5 percent level: a rise in unemployment in one month is associated with a rise in gold in the following month. Unemployment is the only macro variable that is individually significant in this equation, and the next two subsections show that it also appears as the most prominent macro variable in the variance decomposition and impulse responses.

6.2.3 Forecast Error Variance Decomposition

The gold equation tells us which variables matter, but not how much. To rank them, we use forecast error variance decomposition (FEVD), which takes our uncertainty

Table 6.4: Monthly VECM gold equation coefficients. Residual standard error 68.02 on 202 degrees of freedom; adjusted $R^2 = 0.153$; F-statistic 5.741 ($p < 1\text{e}{-6}$).

Coefficient	Estimate	Std. Error	t-value	p-value	Sig.
ect1	−0.0075	0.0080	−0.94	0.349	
ect2	2.4896	0.8624	2.89	0.004	**
Gold (lag 1)	−0.2320	0.0783	−2.96	0.003	**
USD Index (lag 1)	−4.2476	2.8967	−1.47	0.144	
S&P 500 (lag 1)	−0.0748	0.0396	−1.89	0.060	.
Inflation CPI (lag 1)	−9.1361	7.3274	−1.25	0.214	
Treasury 10Y (lag 1)	14.3571	19.3955	0.74	0.460	
Unemployment (lag 1)	14.9939	6.6463	2.26	0.025	*

about where gold will be some months ahead and splits it into the shares contributed by each variable in the system. A variable that explains a large share is an important contributor in the system; one that explains almost none is less important for forecast uncertainty, regardless of whether its coefficient happened to be significant.

Table 6.5 shows how these shares evolve as the horizon lengthens. At one month, gold's own past shocks explain essentially everything, because the other variables have not yet had time to act. As we look further out, their influence grows.

Table 6.5: Monthly FEVD for gold across horizons. Values are percentage shares of gold's forecast error variance attributable to each variable, taken directly from the R output.

Horizon (m)	Gold	USD Idx	S&P 500	Infl. CPI	Treas. 10Y	Unemp.
1	100.00	0.00	0.00	0.00	0.00	0.00
2	97.22	0.05	0.74	0.62	0.01	1.36
3	93.78	0.21	0.63	1.42	0.07	3.90
6	78.83	1.74	0.32	3.92	0.13	15.07
9	65.04	3.08	0.31	5.91	0.14	25.51
12	54.69	3.67	0.54	7.40	0.13	33.57

The shares change with the horizon. At one month gold's forecast error variance is explained almost entirely by its own shocks, echoing the random-walk result of RQ1. By the 12-month horizon the macro variables together account for almost half of gold's forecast error, so they carry more weight once enough time passes. Among them, unemployment has the largest share at 33.57 percent, ahead of CPI at 7.40 percent, the dollar at 3.67 percent, the S&P 500 at 0.54 percent, and the Treasury yield at 0.13 percent. Thus, in this sample and specification, unemployment accounts for a larger share of gold's forecast error variance than CPI, even though CPI is often treated as the more central variable in discussions of gold.

This ranking should be read with its identifying assumptions in mind. The FEVD is Cholesky-identified, which requires an ordering of the variables; the baseline orders them gold, the U.S. Dollar Index, the S&P 500, CPI, the 10-year Treasury yield, and unemployment, so that gold is treated as contemporaneously exogenous and therefore explains 100 percent of its own variance at the first horizon by construction. Because such shares can be sensitive to the ordering, the lag length, and the sample, we re-estimate the decomposition under alternative orderings — placing the macro variables first, placing gold last, and placing unemployment first — and unemployment's share of gold's twelve-month forecast error variance stays between 33.6 and 45.0 percent, the largest macro contribution in every case. The ranking is therefore not an artefact of ordering gold first; if anything, privileging gold less raises unemployment's measured share. We nonetheless read these magnitudes as indicative rather than exact, given the residual diagnostics reported in Section 6.2.6.

A further check concerns the sample period, which the ordering and rank checks above leave open. The 2008–2025 window contains two episodes — the 2008–2009 financial crisis and the 2020 pandemic — in which unemployment and gold both moved sharply, and a variance share estimated over such a window can rest on one of them rather than on a relationship that holds throughout. We therefore re-estimated the twelve-month decomposition on sub-samples that drop each crisis in turn, holding the cointegrating rank and the lag length at their full-sample values so that only the sample changes. Table 6.6 reports the results.

Dropping the 2008–2009 crisis leaves the share essentially unchanged, at 34.4 percent, so the financial crisis is not what produces it. Dropping 2020 is another matter: the share falls to 10.7 percent. On the calm 2010–2019 sub-sample it falls to under 2 percent, and unemployment is no longer the leading macro contributor. A leave-one-year-out check points the same way: 2020 is the only year whose removal pulls the share sharply down, from 33.6 to 10.7 percent. The years whose removal nudges it most in the other direction are themselves crisis years (dropping 2009 alone raises it to 41.7 percent), so the share is keyed to the crisis episodes, and to 2020 above all.

This narrows what the headline figure can be taken to mean. The roughly one-third share belongs to a window that includes the pandemic; it is not a stable feature of the full seventeen years. The reading we carry forward is the weaker one. In this sample gold and unemployment moved together most visibly during the COVID shock, when unemployment reached 14.8 percent and gold passed $2,000 in the same months, and the decomposition is registering that episode more than a relationship that persists across the whole sample. The qualitative result — that unemployment, used as a proxy for broad economic deterioration, outweighs CPI — is robust: it holds in every sub-sample, including those that exclude 2020, and at the conservative rank $r = 1$. Unemployment also stays the single largest macro contributor when any one crisis is removed, dropping 2020 alone included; only in the calm 2010–2019 core, or when several crisis years are excluded together, does another channel (the dollar, the

S&P 500, or the Treasury yield) overtake it. What is specific to 2020 is the magnitude, not the ranking, and we report the ranking on that basis.

Table 6.6: Unemployment's share of gold's twelve-month forecast error variance across sub-samples. The cointegrating rank ($r = 2$) and the lag length are held at their full-sample values; only the estimation sample changes. The final column repeats the exercise at the conservative rank $r = 1$.

Sample	**Months**	**Unemp. (%)**	**CPI (%)**	**Unemp.,** $r = 1$ **(%)**
Full, 2008–2025	212	33.6	7.4	10.7
Excluding 2008–2009	188	34.4	8.3	11.4
Excluding 2020	200	10.7	0.6	4.9
Excluding 2008–09 and 2020	176	6.8	0.2	3.8
2008–2019 (pre-COVID)	144	2.1	0.8	2.1
2010–2019 (calm core)	120	1.7	0.2	0.8

6.2.4 Impulse Response Interpretation

FEVD tells us how much each variable contributes to forecast uncertainty, but not the direction of gold's response. For that we turn to the impulse response functions (IRFs), which trace gold's reaction to a one-time shock in each variable over the following year. The IRFs are computed using the same Cholesky identification and baseline ordering used for the FEVD: gold, the U.S. Dollar Index, the S&P 500, CPI, the 10-year Treasury yield, and unemployment. The responses should therefore be interpreted as conditional on this ordering, and their qualitative interpretation may change under alternative identification choices. Because our monthly sample is relatively short at 212 observations, the 95 percent confidence bands — computed from 1,000 bootstrap replications — are wide, so we read these as qualitative patterns — direction and persistence — rather than precise point estimates.

The most stable response is the response to an unemployment shock (Figure 6.5). It is the only response that is unambiguously signed and stays clear of zero across the entire year: gold rises steadily after an unemployment shock, and even the lower edge of the confidence band stays above zero. This pattern is consistent with a defensive-asset interpretation, in which gold is associated with macroeconomic deterioration, and it aligns with unemployment having the largest macro contribution in the FEVD.

Figure 6.5: Monthly impulse response of gold to an unemployment shock. The response is positive and persistent across the 12-month horizon, and the lower bound of the confidence band remains above zero throughout, suggesting a comparatively robust association in this specification.

The response to a CPI shock (Figure 6.6) runs against the inflation-hedge story: the point estimate is negative throughout and grows in magnitude across the horizon, and while the upper edge of the confidence band sits close to zero, the band as a whole lies below it for most of the year. This is consistent with an inflation shock being associated with a lower gold price on net in this sample. The model itself does not identify why, but one plausible interpretation lies in what accompanied inflation over this period. As described in Chapter 3, the 2022 inflation spike did not occur in isolation: it came alongside the sharpest Federal Reserve rate hikes in decades and a strong dollar. Because gold pays no interest, higher rates make bonds and cash more attractive than gold, and a stronger dollar makes gold costlier for foreign buyers — both would tend to push the price down. Inflation by itself would tend to lift gold, and these accompanying forces may have outweighed it, leaving the net response negative; we present this as an interpretation the data are consistent with, not as a channel the model demonstrates. It is also why a single gold–CPI correlation, of the kind we warned against in Chapter 4, would be so misleading: it cannot separate inflation from the forces that travel with it.

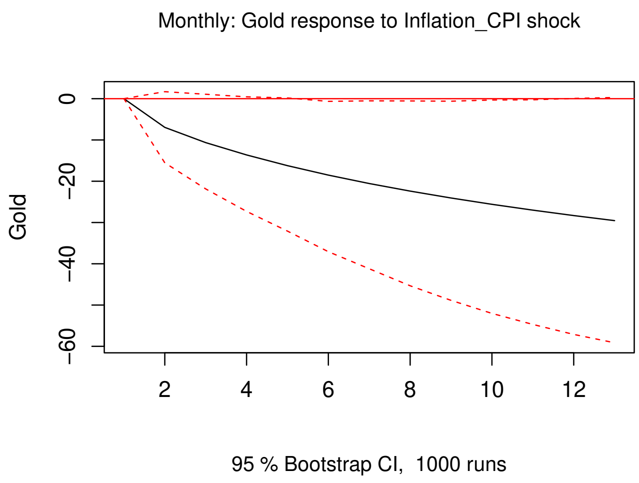

Figure 6.6: Monthly impulse response of gold to an Inflation CPI shock. The point estimate is negative and grows in magnitude across the horizon, with the confidence band lying below zero for most of the year and its upper edge close to zero.

The remaining three responses are minor, and each agrees with its small FEVD share. The dollar response (Figure 6.7) is mildly positive at the monthly horizon. This sits alongside rather than contradicts the negative daily-return correlation between gold and the dollar: the daily figure reflects the immediate trading link, in which a stronger dollar raises gold's price for non-dollar buyers, while the monthly response captures slower joint movements over several months, during which dollar strength and safe-asset demand can coincide.

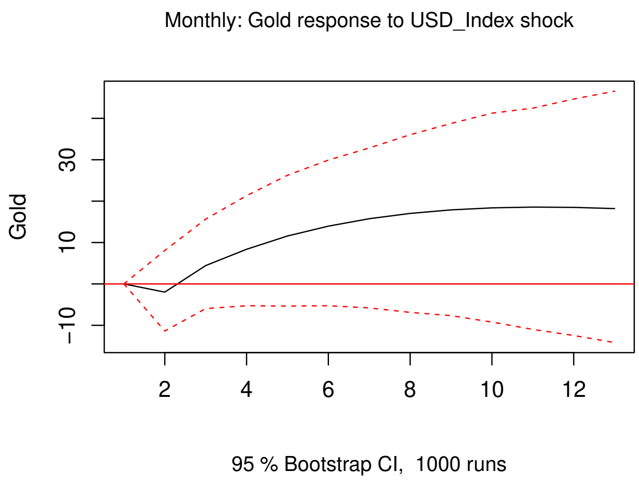

Figure 6.7: Monthly impulse response of gold to a U.S. Dollar Index shock. The point estimate rises modestly and remains positive across the 12-month horizon, although the confidence band crosses zero.

The S&P 500 response (Figure 6.8) is slightly negative at first and turns mildly positive later, but its confidence band straddles zero the whole way, so we read it as essentially no clear effect.

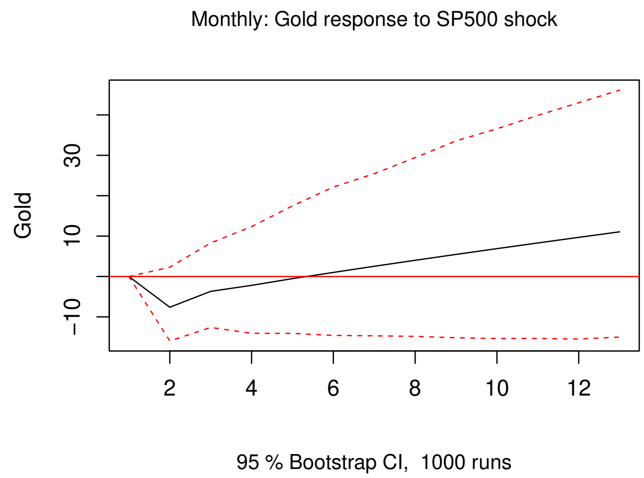

Figure 6.8: Monthly impulse response of gold to an S&P 500 shock. The response is initially slightly negative and turns mildly positive at longer horizons; the confidence band straddles zero throughout.

The Treasury yield response (Figure 6.9) is flat at essentially zero, matching its negligible FEVD share and leaving the yield channel as the weakest of the five.

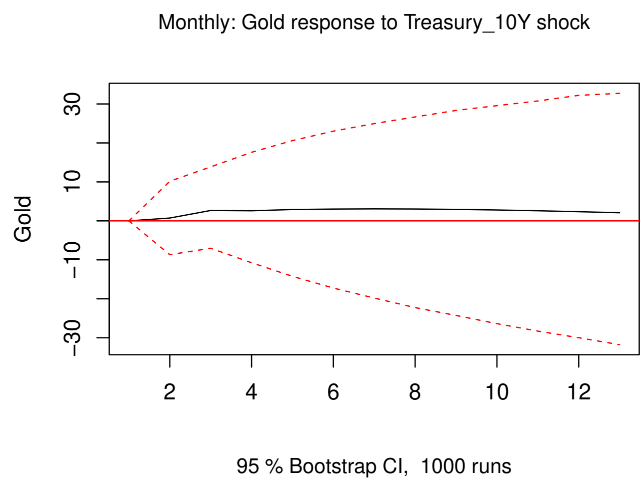

Figure 6.9: Monthly impulse response of gold to a 10-year Treasury yield shock. The point estimate is essentially flat at zero with a wide confidence band.

Across all five, the message is consistent with the variance decomposition, but only the unemployment response supports a stronger, though still conditional, interpretation: its confidence band stays clear of zero, whereas the dollar, S&P 500, and Treasury bands all cross zero. Those three responses should therefore be read as suggestive rather than conclusive. Unemployment provides the most stable signal in this specification; CPI runs opposite to the naive hedge story once its companion forces are present; and the dollar, equities, and yields play only minor and statistically weak roles.

6.2.5 Granger Causality

As a final, stricter check, we ask whether each variable helps predict gold in the short run — the question a Granger causality test answers. Because our series are integrated of order one, the usual test is invalid, so we use the lag-augmented procedure of Toda and Yamamoto (1995), which fits the model in levels with one extra lag and tests only the original lags, restoring a valid test without having to difference the data and lose the long-run information.

On this test, none of the five variables Granger-causes gold at the 5 percent level; unemployment comes closest ($p \approx 0.06$) but does not cross the threshold, while the S&P 500, the dollar, CPI, and the Treasury yield are clearly insignificant. This result deserves weight in its own right: the evidence for short-run, month-to-month predictability of gold from these variables is weak, and unemployment, though it stands out in the FEVD and the impulse responses, does not pass the stricter Granger test. The contrast is not a contradiction. Granger causality looks for tight, month-to-month predictability, whereas the FEVD and the impulse responses capture effects that accumulate gradually over a year — precisely the kind of slow transmission a short-run test is least able to detect. The consistent thread is that unemployment appears most prominently across the three methods, even though the strength of the evidence varies with the test.

6.2.6 Diagnostics and Caveats

The residual diagnostics behave as one would expect for a macro-financial model estimated over a crisis-rich sample, and they call for caution in reading the results. The Portmanteau test rejects the absence of serial correlation ($p = 0.007$); the Jarque–Bera test rejects normality, reflecting the fat tails and large outliers around 2008–2009 and 2020; and the multivariate ARCH test rejects, indicating that volatility clusters within the system. Taken together, these failures mean the VECM estimates should be treated as indicative rather than definitive, and that formal inference on individual coefficients is unreliable. Our reading rests not on precise coefficient values or exact shares, but on the broad, repeated pattern across the FEVD, the impulse responses, and the Granger test — the repeated prominence of unemployment and the relative weakness of the other channels — rather than on any single estimate.

We also check that the relationships are stable over time, rather than an artefact of one sub-period, using the OLS-CUSUM test for each equation of the monthly system (Figure 6.10). None of the test paths crosses the 5 percent bounds, which supports treating the estimated relationships as stable across the whole sample. The near-unit root that appears in the lag-augmented levels system used for the Toda–Yamamoto test reflects the I(1) and cointegrated structure of the data and is expected rather than a sign of misspecification; parameter stability is assessed by the OLS-CUSUM paths, not by that root.

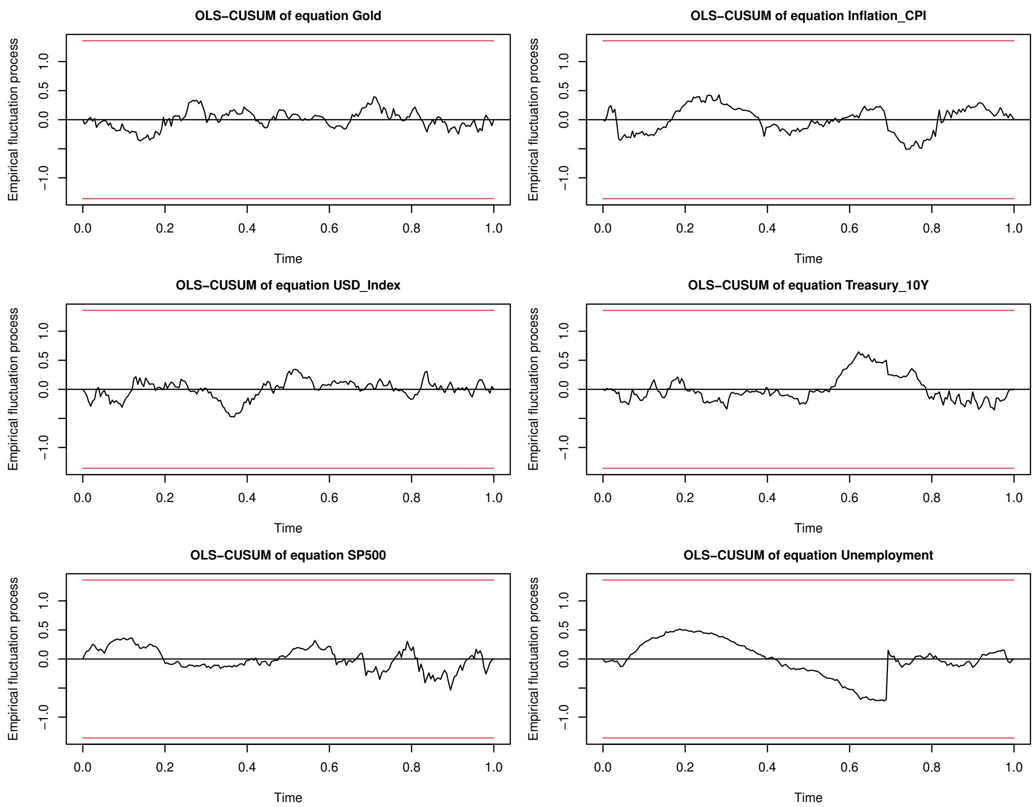

Figure 6.10: OLS-CUSUM stability tests for each of the six equations of the monthly system. None of the empirical fluctuation processes crosses the 5% significance bounds, supporting parameter stability across the sample.

Across the three methods, RQ2 revises the conventional picture of gold. Gold reacts to the dollar and to yields in the expected directions but only weakly; its reduced-form response to inflation runs in the opposite direction to the simple hedge story once the tightening and dollar channels are present in the sample; and the variable associated with the largest share of its forecast error variance is the unemployment rate. We read unemployment here as a proxy for broad macroeconomic deterioration rather than as a force that mechanically moves gold: it is a lagging business-cycle indicator, not a direct measure of uncertainty or financial stress, and the association we document is consistent with investors moving into gold as the economy weakens, not with unemployment mechanically causing the gold price to rise. On this reading gold behaves more like a conditional hedge against economic deterioration than like an inflation hedge in the narrow sense, a characterization that the diagnostics ask us to treat as indicative rather than definitive.

6.3 RQ3: Comparing Bitcoin and Gold

The behavior of gold and Bitcoin during the March 2020 crisis previews the central result of this section. In that month the S&P 500 fell about 12 percent on a month-end basis and roughly 34 percent peak-to-trough, from its February 19 high to its March 23 low; gold ended the month slightly higher after a sharp mid-month dash-for-cash dip, while Bitcoin declined alongside equities. The remainder of this section examines whether the full sample confirms this pattern.

The "digital gold" label was applied to Bitcoin around 2017. The argument rested on shared institutional features: Bitcoin's supply is capped at 21 million coins, no central bank can expand it, and it operates outside the traditional banking system, all properties that gold also possesses. If scarcity, statelessness, and costly production are the sources of gold's defensive value, Bitcoin might be expected to perform the same function during periods of market distress.

Two forms of evidence bear on this claim. The first is the volatility of each asset's price. The second is the correlation of each asset with equities, since the defensive value of gold derives from its tendency to rise when equities fall. Because Bitcoin entered our dataset in September 2014, this comparison is conducted over the period 2014–2025.

6.3.1 Volatility Comparison

Table 6.7: Annualized 30-day rolling volatility, 2014–2025.

Asset	Mean	Median	Min	Max
Gold	14.39%	13.78%	6.86%	42.05%
S&P 500	15.26%	12.59%	3.57%	85.31%
Bitcoin	62.17%	58.76%	14.57%	153.52%

Gold's mean annualized volatility is about 14 percent, and that of the S&P 500 is about 15 percent, while Bitcoin's is 62 percent, roughly four times higher than either. The extremes are equally pronounced. In its least volatile 30-day window, Bitcoin's annualized volatility was 14.6 percent, comparable to gold's average; in its most volatile window, it reached 154 percent, whereas gold's maximum was 42 percent. Figure 6.11 displays the three series, in which the Bitcoin line lies on a markedly larger scale than the other two for nearly the entire sample.

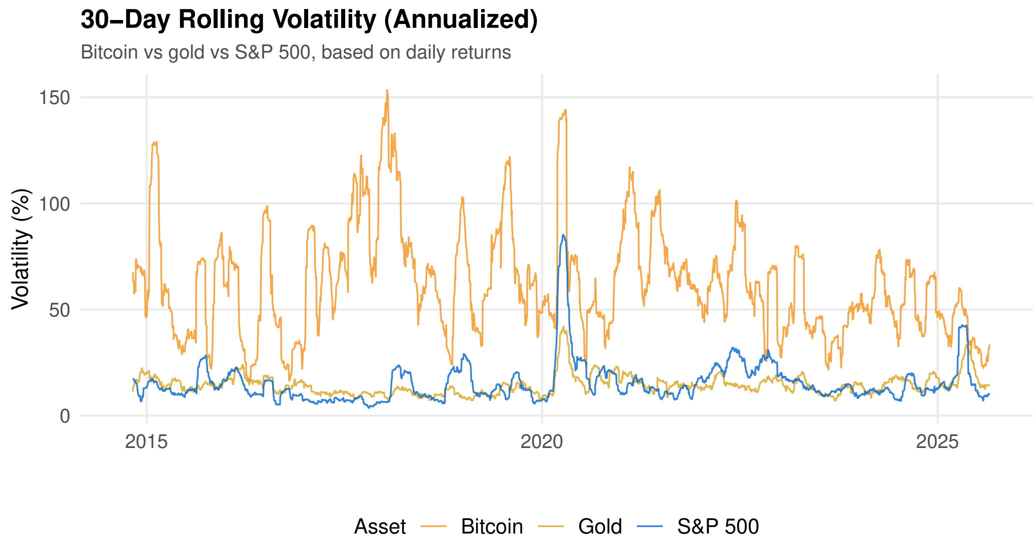

Figure 6.11: 30-day rolling volatility, annualized, for Bitcoin, gold, and the S&P 500, 2014–2025. Bitcoin's series sits visibly above gold and the S&P 500 across the entire sample.

A Levene test on the dispersion of daily returns rejects the null of equal volatility across the three assets (the more familiar F-test, which assumes normality, agrees but is less reliable given the fat tails of daily returns). A GARCH(1,1) model with Student-t errors, which passes the diagnostic checks, confirms this result: Bitcoin's conditional volatility is several times that of gold across the sample. The news impact curves in Figure 6.12 show that even the conditional response to a small standardized shock is substantially larger for Bitcoin than for gold or the S&P 500. On the volatility criterion, Bitcoin and gold are not comparable; Bitcoin has been substantially more volatile throughout the sample.

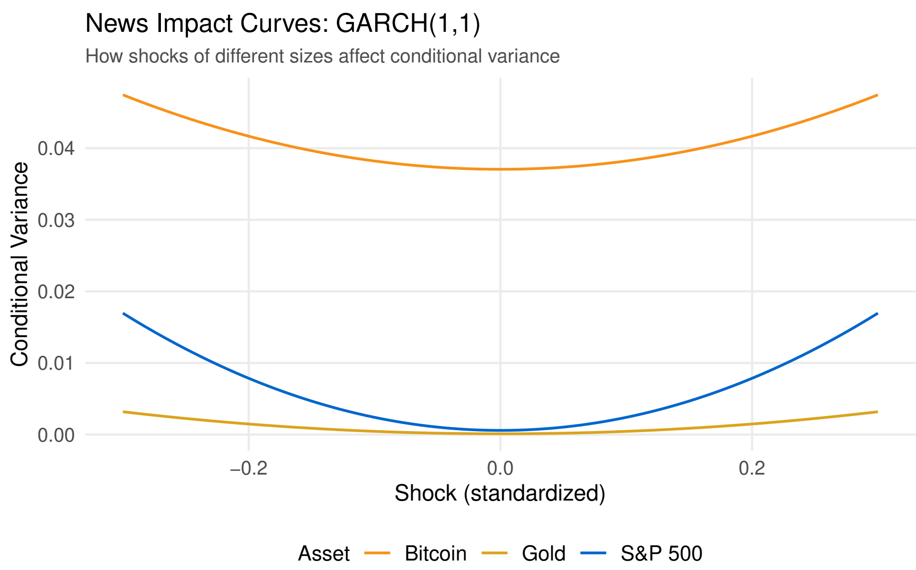

Figure 6.12: GARCH(1,1) news impact curves with Student-t errors. The Bitcoin curve sits well above the gold and S&P 500 curves at every shock size, reflecting a fundamentally higher conditional volatility regime.

The same conclusion is visible when we plot the fitted conditional volatilities directly. Figure 6.13 shows the GARCH(1,1) conditional standard deviation of each asset over the 2014–2025 window. Bitcoin's volatility series sits on a markedly higher level than gold and the S&P 500 throughout, and its spikes during stress episodes are correspondingly larger. Gold remains the lowest and most stable of the three, as we expect from a defensive asset.

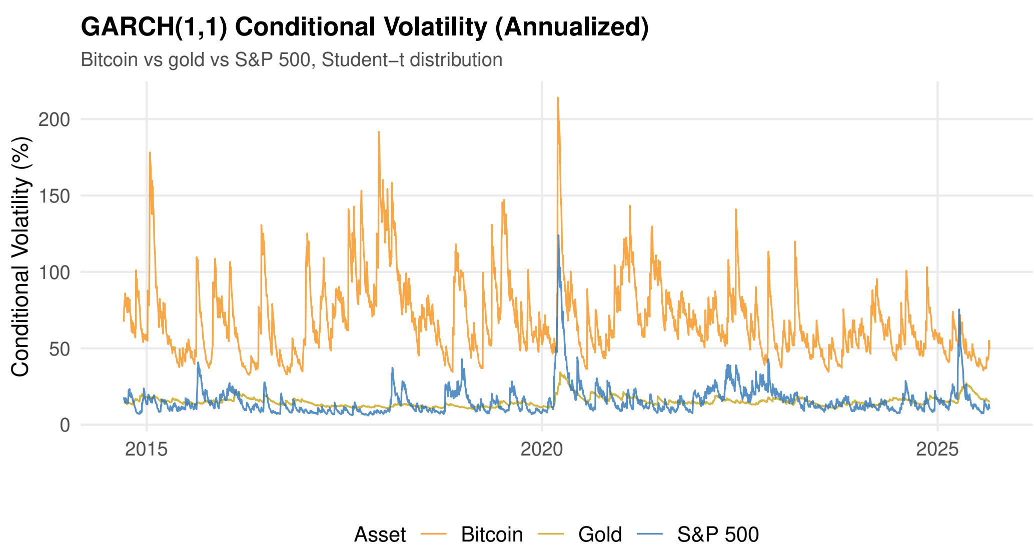

Figure 6.13: GARCH(1,1) conditional volatility for Bitcoin, gold, and the S&P 500, 2014–2025. Bitcoin's conditional volatility is several times that of gold across the entire sample, while gold remains the lowest and most stable series.

Before we rely on these volatility estimates, we check that the fitted GARCH(1,1) models are well specified. The standardized-residual diagnostics for gold, Bitcoin, and the S&P 500 are collected in Appendix A (Figures A.1 through A.3). In each case the standardized residuals fluctuate around zero without obvious structure, the histogram is approximately symmetric, the Q–Q plot against the Student-t reference is close to linear through the body of the distribution, and the autocorrelation function of the squared standardized residuals shows no remaining significant lags. The absence of leftover autocorrelation in the squared residuals is the key result: it indicates that the GARCH(1,1) specification has absorbed the volatility clustering in each return series, so the conditional-volatility comparison above rests on adequately specified models.

6.3.2 Correlation with Equities

The second question is whether Bitcoin moves with or against equities. A hedge is uncorrelated with or moves opposite to equities when they decline, whereas a risk asset declines with them. Table 6.8 reports the DCC-GARCH results for the 2014–2025 period.

Table 6.8: DCC-GARCH dynamic correlation summary, 2014–2025.

Pair	**Mean**	**Median**	**Min**	**Max**	**% Time Neg.**
Gold ↔ S&P 500	−0.0217	−0.0168	−0.4384	0.3340	56.6%
Bitcoin ↔ S&P 500	0.1622	0.1468	−0.2189	0.5038	8.1%
Gold ↔ Bitcoin	0.0775	0.0714	−0.2037	0.3920	21.0%

Gold's mean conditional correlation with the S&P 500 is −0.02 (median −0.02), and the correlation is negative on 56.6 percent of trading days; this is consistent with hedge behavior. Bitcoin's mean conditional correlation with the S&P 500 is +0.16: small but clearly positive, and negative on only 8.1 percent of trading days, which implies positive co-movement with equities on roughly 92 percent of days. This is consistent with a risk asset rather than a hedge. The correlation between gold and Bitcoin itself averages +0.08, well below the near-unity correlation that the digital-gold label would imply.

Figure 6.14 plots the two equity-linked correlation series over time: gold's series remains near zero throughout, while Bitcoin spends most of the post-2020 sample in clearly positive territory.

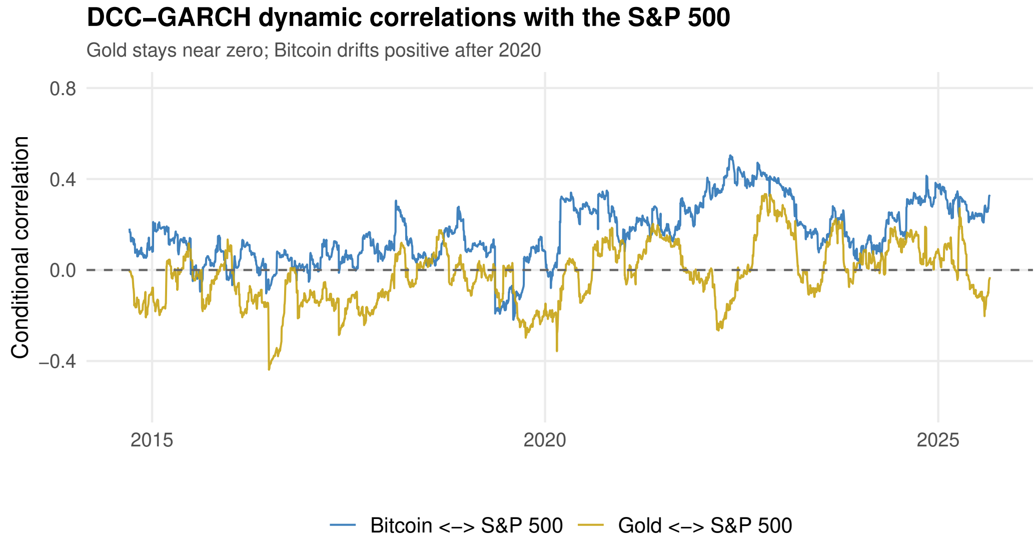

Figure 6.14: DCC-GARCH dynamic conditional correlations with the S&P 500. Gold's correlation hovers near zero and dips into negative territory through stress episodes; Bitcoin's correlation drifts upward and stays mostly positive after 2020.

Figure 6.15 shows the gold–Bitcoin correlation, which is small and positive for most of the sample, with no evidence of the near-unity relationship implied by the digital-gold label.

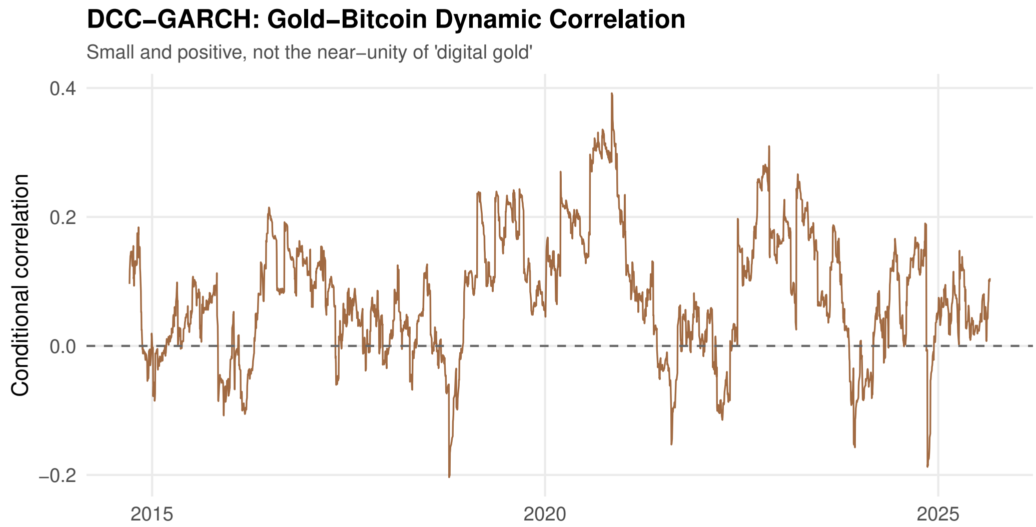

Figure 6.15: DCC-GARCH dynamic correlation between gold and Bitcoin. The series spends most of its time positive but small, peaking around 2020–2021 and crossing into negative territory only briefly.

As a robustness check, we compute a non-parametric 90-day rolling correlation (Figure 6.16), which closely tracks the DCC-GARCH series; the result therefore does not depend on the choice of model.

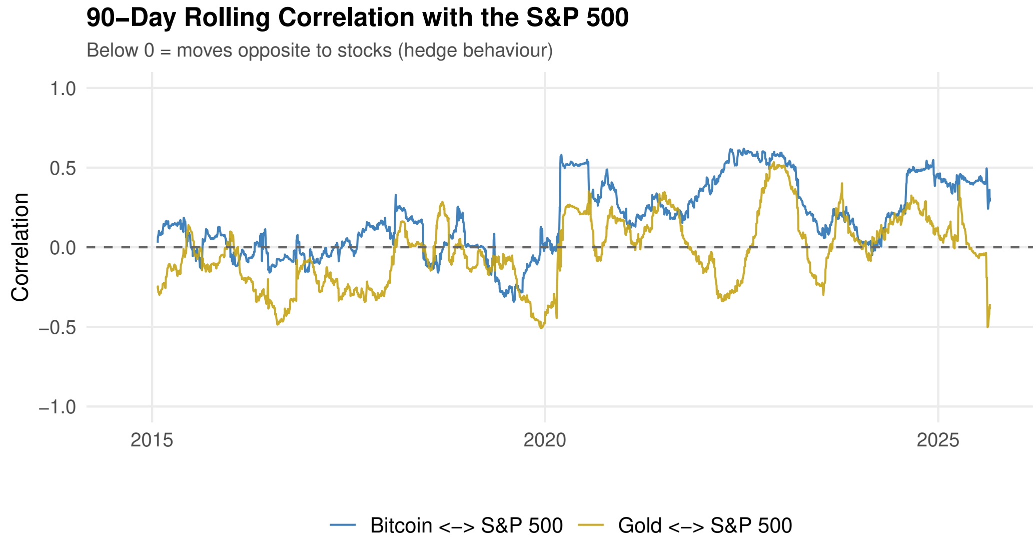

Figure 6.16: 90-day rolling correlation with the S&P 500 for Bitcoin and gold. The rolling-window evidence agrees with the DCC-GARCH series: Gold drifts around zero, and Bitcoin moves into clearly positive territory after 2020.

6.3.3 Year-by-Year Correlation and the Post-2020 Shift

The full-sample averages are informative, but the annual decomposition uncovers a structural change. From 2014 through 2019, Bitcoin's annual correlation with the S&P 500 was small and mostly low: it was negative in several years (for example -0.09 in 2014, -0.04 in 2016, and -0.12 in 2019) and only weakly positive in others, while gold's annual correlations were of comparable magnitude. During these years there was a defensible case for the digital-gold characterization; Bitcoin was far more volatile but remained largely independent of equities.

The annual volatility pattern is shown in Figure 6.17. Bitcoin's annualized volatility exceeds that of gold and the S&P 500 in every year of the sample, including its least volatile years.

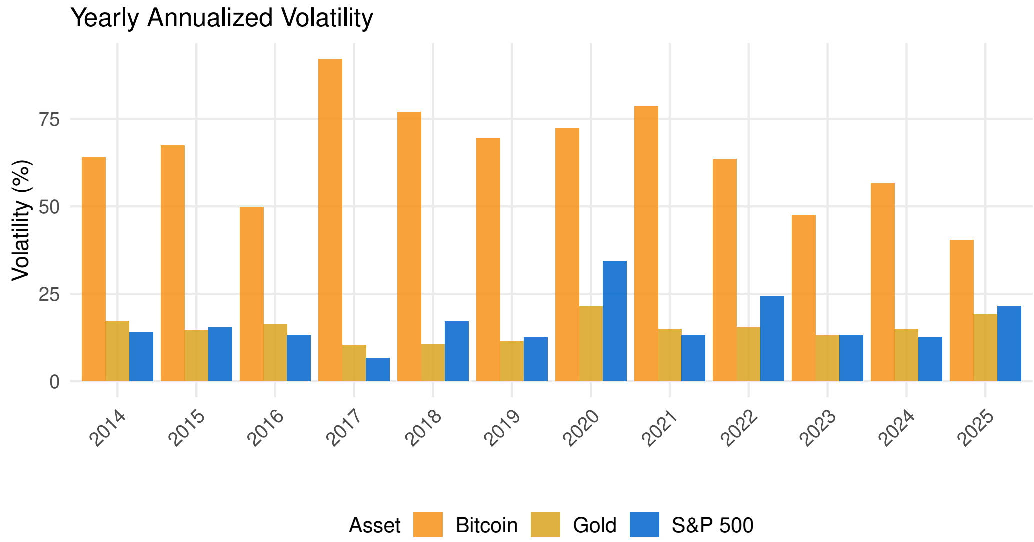

Figure 6.17: Annualized yearly volatility, 2014–2025. Bitcoin's bars dwarf gold and the S&P 500 every year of the sample.

In March 2020, the S&P 500 fell roughly a third from its February peak, and Bitcoin declined with it. The correlation pattern did not subsequently revert. Table 6.9 reports the annual correlations.

Table 6.9: Annual correlation with the S&P 500.

Year	Bitcoin ↔ S&P 500	Gold ↔ S&P 500
2014	−0.09	−0.21
2019	−0.12	−0.27
2020	+0.43	+0.18
2021	+0.26	+0.06
2022	+0.56	+0.11
2024	+0.36	+0.17
2025	+0.40	−0.03

Bitcoin's correlation with the S&P 500 changes sign and remains positive from 2020 onward, whereas gold's correlation remains near zero throughout. Figure 6.18 displays the same series. The timing of the shift is notable: the years in which Bitcoin moved from a niche asset toward mainstream adoption, through ETF approvals, corporate treasury allocations, and institutional participation, are the same years in which its behavior moved toward equities and away from gold. Increased institutional integration is associated with weaker, not stronger, support for the digital gold characterization.

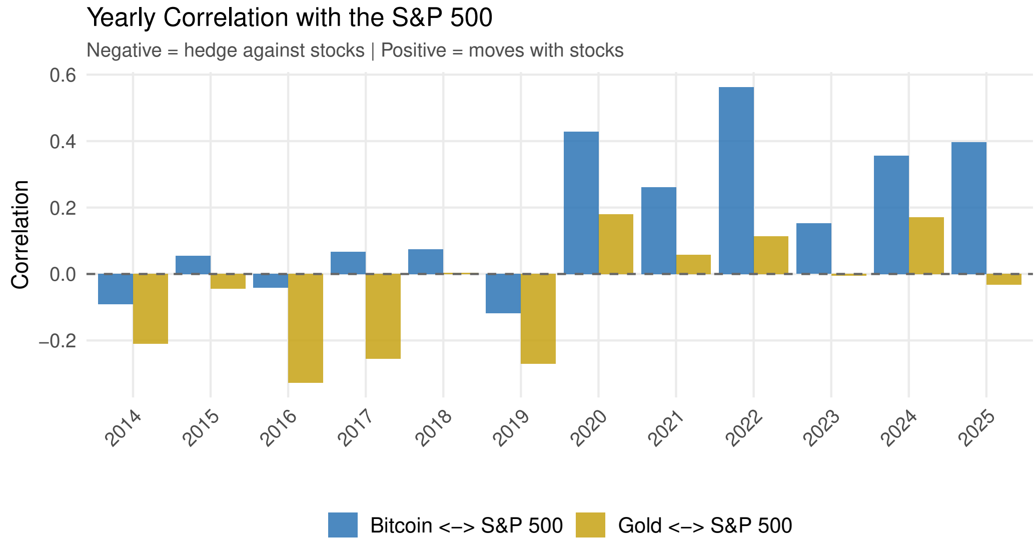

Figure 6.18: Yearly correlation with the S&P 500. Gold (amber bars) sits near zero with a slight negative tilt before 2020; Bitcoin (blue bars) flips from near zero to clearly positive starting in 2020 and stays there.

6.3.4 Behaviour Across VIX Stress Regimes

The annual decomposition shows when the correlation structure changed; a complementary question is how gold and Bitcoin behave specifically on days of elevated market stress, which is the defining test of a safe haven. A hedge need only be uncorrelated with equities on average, but a safe haven must hold up during the worst episodes (Baur and Lucey, 2010). We therefore split the daily sample by the level of the official CBOE VIX. Following the quantile-based approach common in the flight-to-safety literature (Löwen et al., 2021), which judges safe-haven status by behaviour in the extreme tail of the stress distribution rather than at average stress, we classify a trading day as high-stress when the VIX exceeds its 75th sample percentile (a level of 21.25) and as calm otherwise. This places 688 of the 2,753 trading days in the high-stress regime. As a robustness check on the most extreme conditions, we also report a stricter split at the 90th percentile (a VIX level of 26.90), which isolates the most stressed roughly ten percent of days. One limitation of this stress measure should be stated: the VIX captures expected volatility in the U.S. equity market, so the regime classification is appropriate for a U.S.-centred analysis but may not capture every form of global macroeconomic or geopolitical stress, some of which need not register in U.S. equity-option prices.

Table 6.10 reports each asset's correlation with the S&P 500 within the two regimes, and the result departs from the textbook safe-haven prediction. Gold's correlation with the S&P 500 is slightly negative in calm periods (-0.037) but turns mildly positive in high-stress periods ($+0.073$), and rises further to $+0.163$ on the most extreme top-decile days. Strictly interpreted, this runs against the classical safe-haven definition, under which gold's equity correlation should fall, not rise, when markets are stressed. The pattern is consistent with the well-documented observation that in periods of extreme volatility even defensive assets can briefly co-move with equities, as investors liquidate whatever they can in a broad "dash for cash" (King and Wadhwani, 1990). We also note a statistical caveat: correlations computed on high-volatility subsamples are mechanically biased upward relative to the full sample (Forbes and Rigobon, 2002), so part of the rise for both assets is an artefact of conditioning on stress; the decisive evidence is therefore the comparison between the two assets within the same regime, not the level of either correlation against an absolute benchmark.

Table 6.10: Correlation with the S&P 500 by VIX stress regime. High-stress days are those on which the official CBOE VIX exceeds its 75th sample percentile (21.25); calm days are the remainder. The lower panel reports the stricter top-decile cut (VIX above 26.90).

Regime	**Gold ↔ S&P 500**	**Bitcoin ↔ S&P 500**
Primary split (top quartile)		
Low VIX (calm)	−0.037	0.067
High VIX (stress)	0.073	0.393
Robustness (top decile)		
Rest	−0.054	0.110
Extreme VIX (top 10%)	0.163	0.488

The comparison that matters is between the two assets, not against an idealized benchmark. On both criteria gold remains far more defensive than Bitcoin. Gold's stress-period correlation of +0.07 (and +0.16 at the extreme) is modest, whereas Bitcoin's correlation with the S&P 500 jumps from +0.07 in calm periods to +0.39 under stress, and to +0.49 on the most extreme days, roughly a six-fold increase. Bitcoin moves most with equities precisely when an investor seeking protection would least want it to, the behaviour of a risk asset. Gold's correlation rises only modestly and from a near-zero base, so although it does not satisfy the strict safe-haven definition in our sample, it keeps a substantially more defensive profile than Bitcoin throughout. The regime evidence agrees with the rest of RQ3: relative to Bitcoin, gold is the far more defensive of the two, while Bitcoin's sharp rise in equity co-movement under stress is strong evidence against the digital-gold characterization.

6.3.5 Gold's Behavior Across the Sample

Over the same period, gold's behavior was consistent with its historical role. It rose during the 2008 crisis as equities declined; it exceeded $2,000 per ounce for the first time in August 2020, when unemployment reached 14.8 percent; and it held its value through the 2022 inflation shock and the rapid tightening cycle before rising above \$3,400 in 2024–2025. Low volatility, a low average correlation with equities, and a tendency to appreciate during periods of distress characterized gold across the sample. As the regime split shows, this correlation is not strictly negative on the most stressed days, so gold does not meet the narrowest safe-haven definition; but it remains far more defensive than Bitcoin, consistent with the broader role of a hedge.

The digital gold claim posed a reasonable question when it was first raised. The evidence in this sample indicates that Bitcoin is approximately four times more volatile than gold, that it moves with equities rather than against them, and that this co-movement has strengthened over time, whereas gold has retained a more defensive profile on average. The evidence, therefore, supports describing gold as a conditional hedge and Bitcoin as a speculative asset.

6.4 Putting the Three Questions Together

Taken together, the three research questions yield a consistent picture of gold. One day ahead, gold is essentially a random walk: the most recent price forecasts tomorrow but says nothing about next quarter. Over longer horizons gold's own shocks still explain most of its forecast error variance, while unemployment is the macro variable that appears most consistently in the FEVD and in the most persistent impulse response; the size of that variance share, though, owes much to the 2020 pandemic, so it is the ranking and not the magnitude that we lean on. The dollar and the 10-year yield play secondary roles, and the CPI response works against the simple inflation-hedge story once the tightening and dollar channels are taken into account. On volatility and equity correlation, gold behaves as a conditional hedge across the sample — low volatility and a near-zero average correlation with equities, though not a strictly negative correlation on the most stressed days — whereas Bitcoin behaves as a speculative asset, a contrast that has grown sharper since 2020.

No single model settles what gold and Bitcoin are doing in an investor's portfolio. The strength of the analysis is that, applied to one seventeen-year dataset, ARIMA, the VECM, GARCH(1,1), and DCC-GARCH agree, and several results, especially the unemployment ranking, appear only when the methods are read together and interpreted with the stated caveats.

CHAPTER 7: CONCLUSION

7.1 Main Conclusions

This thesis analysed gold, Bitcoin, and macroeconomic dynamics across a crisis-rich sample from 2008 to 2025, asking whether historical gold prices forecast future prices, whether macroeconomic variables explain gold's dynamics, and whether Bitcoin behaves like gold or like a speculative risk asset.

Gold behaves like a random walk in the short run. One-step-ahead forecasts can be accurate, but only because the most recent price is the best immediate guide; forecasts at longer horizons based on gold's history alone have little power.

Gold's macroeconomic behaviour, in turn, appears tied more closely to broad economic deterioration than to inflation alone. The monthly VECM shows that unemployment shocks are associated with a positive response in gold and with the largest share, about a third, of its forecast error variance at the twelve-month horizon. That share is concentrated in the 2020 pandemic — it falls to about a tenth once 2020 is excluded — so what we rely on is the ranking of unemployment above the other macro variables and the sign of its effect, not the exact size; we read unemployment as a proxy for macroeconomic deterioration rather than as a mechanical cause of the gold price. The CPI response runs counter to the textbook inflation-hedge story: the point estimate is negative, even though consumer prices and gold are highly correlated in levels over the long run. One plausible interpretation, consistent with the data but not demonstrated by the model, is that this reflects the net effect of inflation shocks together with the monetary tightening, real-yield, and dollar movements that accompanied them in our sample, rather than inflation itself depressing gold. On this reading gold behaves more like a conditional hedge against economic deterioration than like a simple inflation hedge, a characterization we treat as indicative given the VECM diagnostics.

Bitcoin, finally, should not be treated as digital gold in the safe-haven sense. It is far more volatile than gold and grows more positively correlated with the S&P 500 after 2020, so while it may serve speculative or return-seeking purposes, it does not provide gold's defensive function. Gold's own defensiveness, though, needs stating carefully: the VIX stress-regime split shows its equity correlation rising modestly on the most stressed days rather than turning more negative, so gold does not satisfy the strictest safe-haven definition either. Its defensiveness is relative, far stronger than Bitcoin's and anchored in low volatility and a near-zero average correlation, rather than a guarantee of negative co-movement in every crisis.

7.2 Contribution of the Thesis

The broader contribution is to show that asset labels have to be tested empirically. Gold is not automatically a safe haven in every period, and Bitcoin is not a substitute for gold simply because its supply is limited. Separating the two requires forecasting models, macroeconomic linkages, volatility analysis, and dynamic correlations together; combined, they support a clear distinction. Gold behaves as a conditional hedge and defensive asset, steadier than Bitcoin and near-uncorrelated with equities on average, though not negatively correlated on the most stressed days, while Bitcoin behaves primarily as a speculative risk asset.

The expanded literature review ties each empirical method to a research tradition: ARIMA to the random-walk problem in financial forecasting, the VECM to long-run macro-financial transmission (with the monthly frequency following the standard point that cointegration analysis should use the natural release frequency of the underlying variables), GARCH and DCC-GARCH to volatility clustering and time-varying dependence, and the official CBOE VIX to market fear and expected volatility rather than treating it as a merely descriptive variable.

7.3 Comparison with Expectations

Table 7.1: Expected versus observed findings.

Expectation	**Observed result**
Gold is a simple short-run inflation hedge	Not supported; the monthly IRF for the CPI shock is negative, although CPI and gold are highly related in levels over the long run
Gold is sensitive to dollar movements	Daily return correlation is negative, as expected; the monthly IRF is mildly positive with a wide confidence band
Gold is associated with broad economic deterioration	Supported in this sample, with caveats; unemployment has the largest macro FEVD share and the most persistent positive IRF, but that share is driven mainly by the 2020 pandemic and it does not pass the 5% Granger-causality test
Bitcoin acts as digital gold during stress	Not supported in the post-2020 period; correlation with equities rises sharply after 2020
Bitcoin's volatility resembles gold's	Not supported; Bitcoin volatility is roughly four times higher than gold's on average

7.4 Limitations and Future Research

Our thesis has several limitations. The sample is dominated by U.S. macroeconomic variables and the U.S. dollar; extending it to euro-area or emerging-market data would test whether the conditional-hedge interpretation generalizes. We restricted the monthly VECM to six variables to preserve degrees of freedom; future work could add policy-related variables such as central bank balance sheets or geopolitical risk indices. Bitcoin's post-2020 behavior rests on a relatively short history; we will need longer samples to determine whether the equity-like correlation we observe since 2020 is a permanent feature or a phase tied to a specific monetary policy environment. Finally, GARCH-family models capture conditional volatility but not all forms of tail risk; extensions to extreme-value methods or copula-based models could refine the safe-haven interpretation during crisis episodes.

CHAPTER A: GARCH DIAGNOSTIC FIGURES

This appendix collects the standardized-residual diagnostics for the GARCH(1,1) models fitted to gold, Bitcoin, and the S&P 500 in Section 6.3.1.

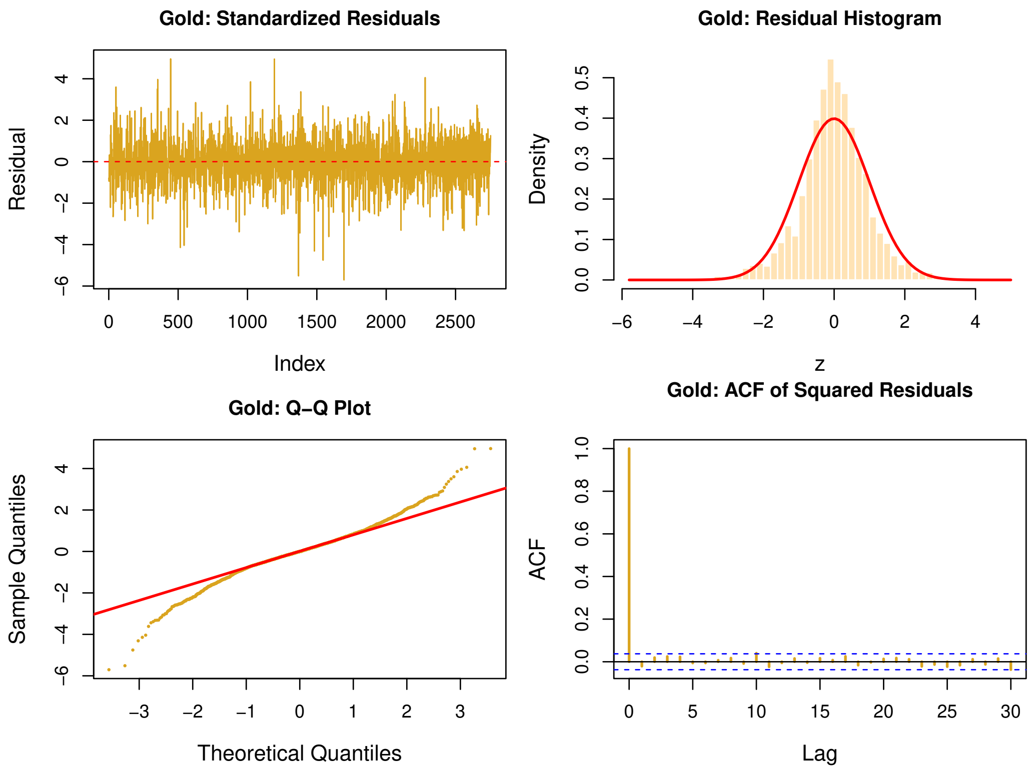

Figure A.1: GARCH(1,1) standardized-residual diagnostics for gold. The residuals show no remaining structure and the squared-residual ACF lies within the confidence bands, indicating the volatility clustering has been adequately captured.

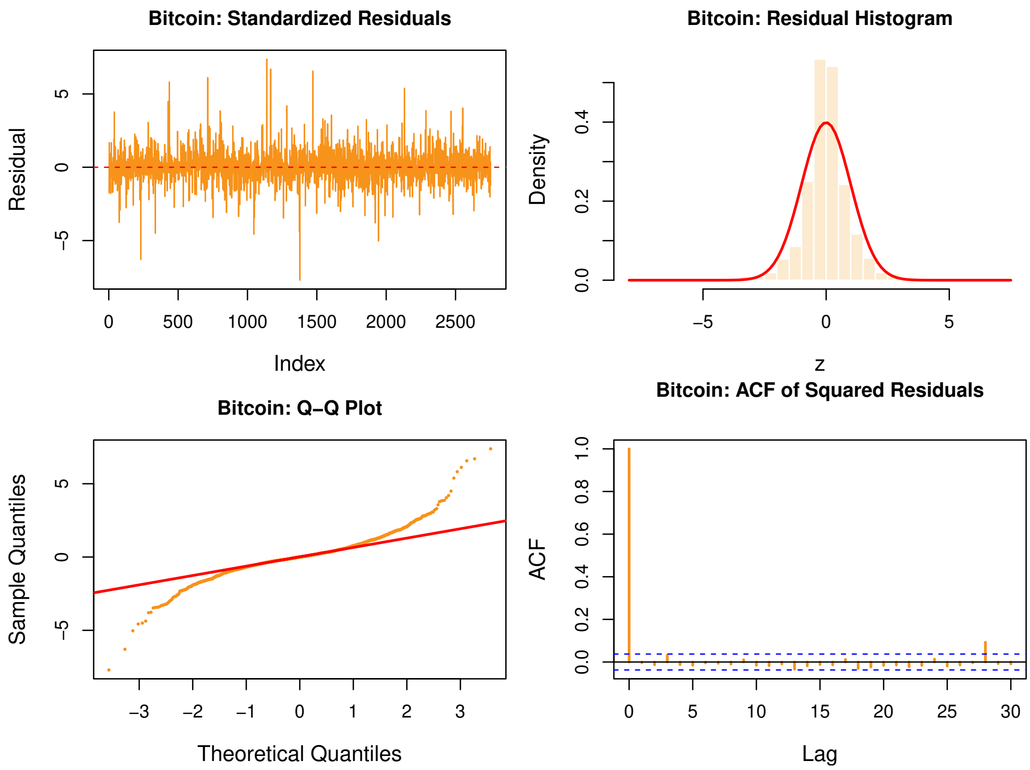

Figure A.2: GARCH(1,1) standardized-residual diagnostics for Bitcoin. As with gold, the squared-residual ACF shows no significant remaining lags, supporting the specification despite Bitcoin's much higher volatility level.

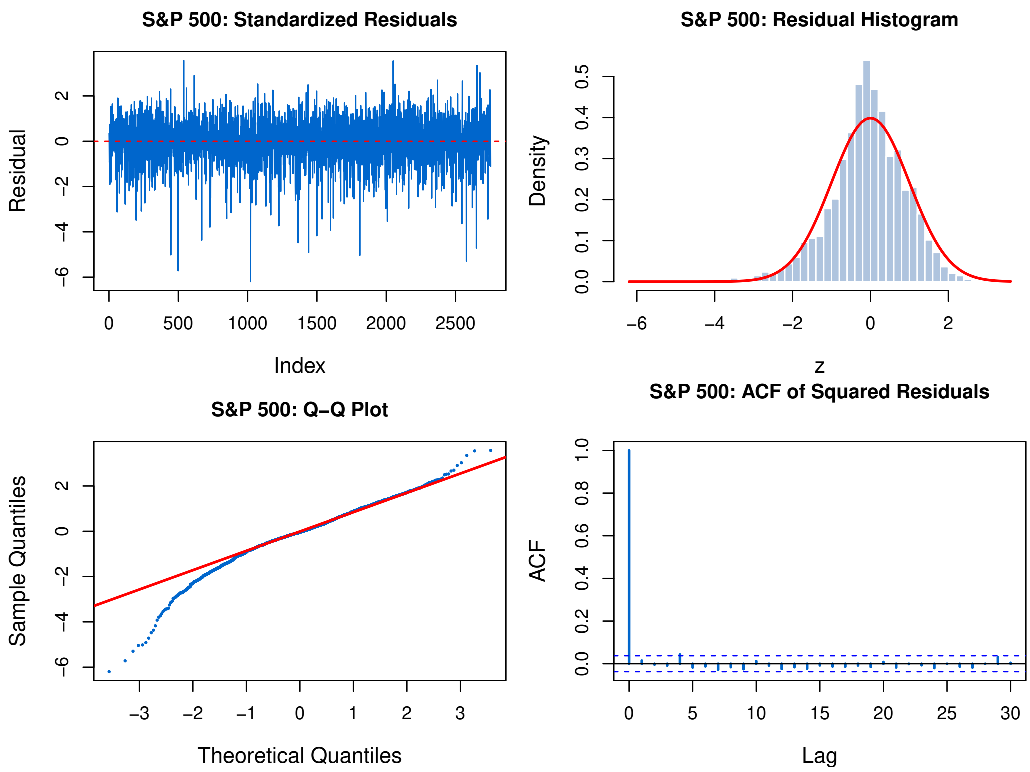

Figure A.3: GARCH(1,1) standardized-residual diagnostics for the S&P 500. The diagnostics again indicate an adequately specified conditional variance model.

BIBLIOGRAPHY

Acikgoz, T. (2025). Gold and Bitcoin as hedgers and safe havens: Perspective from nonlinear dynamics. Resources Policy, 102, 105489. https://doi.org/10.1016/j.resourpol.2025.105489

Al-Nassar, N. S. (2024). Can gold hedge against inflation in the UAE? A nonlinear ARDL analysis in the presence of structural breaks. PSU Research Review, 8(1), 151–166. https://doi.org/10.1108/PRR-08-2021-0038

Baur, D. G., and Lucey, B. M. (2010). Is gold a hedge or a safe haven? An analysis of stocks, bonds, and gold. Financial Review, 45(2), 217–229.

Baur, D. G., and McDermott, T. K. (2010). Is gold a safe haven? International evidence. Journal of Banking & Finance, 34(8), 1886–1898.

Ben Jabeur, S., Mefteh-Wali, S., and Viviani, J.-L. (2024). Forecasting gold price with the XGBoost algorithm and SHAP interaction values. Annals of Operations Research, 334, 679–699. https://doi.org/10.1007/s10479-021-04187-w

Bouri, E., Molnar, P., Azzi, G., Roubaud, D., and Hagfors, L. I. (2017). On the hedge and safe haven properties of Bitcoin: Is it really more than a diversifier? Finance Research Letters, 20, 192–198.

Box, G. E. P., and Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control. Holden-Day.

Cboe Global Markets (2024). Cboe Volatility Index (VIX) Methodology. Available at: https://cdn.cboe.com/resources/vix/VIX_Methodology.pdf

Dyhrberg, A. H. (2016). Bitcoin, gold, and the dollar: A GARCH volatility analysis. Finance Research Letters, 16, 85–92.

Engle, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. Journal of Business & Economic Statistics, 20(3), 339–350.

Forbes, K. J., and Rigobon, R. (2002). No contagion, only interdependence: Measuring stock market comovements. The Journal of Finance, 57(5), 2223–2261.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econometrica, 59(6), 1551–1580.

Joy, M. (2011). Gold and the US dollar: Hedge or haven? Finance Research Letters, 8(3), 120–131.

King, M. A., and Wadhwani, S. (1990). Transmission of volatility between stock markets. The Review of Financial Studies, 3(1), 5–33.

Kumar, A. S., Mohan, M., and Niveditha, P. S. (2025). Assessing Bitcoin and gold as safe havens amid global uncertainties: A rolling window DCC-GARCH analysis. Global Business Review (advance online publication). https://doi.org/10.1177/09711023251322578

Löwen, C., Kchouri, B., and Lehnert, T. (2021). Is this time really different? Flight-to-safety and the COVID-19 crisis. PLOS ONE, 16(5), e0251752.

Snene Manzli, Y., and Jeribi, A. (2024). Shelter in uncertainty: Evaluating gold and Bitcoin as safe havens against G7 stock market indices during global crises. Scientific Annals of Economics and Business, 71(3), 417–447. https://doi.org/10.47743/saeb-2024-0011

Snene Manzli, Y., Fakhfekh, M., Béjaoui, A., Alnafisah, H., and Jeribi, A. (2025). On the hedge and safe-haven abilities of bitcoin and gold against blue economy and green finance assets during global crises: Evidence from the DCC, ADCC and GO-GARCH models. PLOS ONE, 20(2), e0317735. https://doi.org/10.1371/journal.pone.0317735

Nguyen, T. T. B. (2024). Comparative analysis of gold, art, and wheat as inflation hedges. Journal of Risk and Financial Management, 17(7), 270. https://doi.org/10.3390/jrfm17070270

Pflaumer, P. (2024). Gold price dynamics: Geometric random walk and ARIMAX models. Working paper, TU Dortmund University (Eldorado repository). https://eldorado.tu-dortmund.de/bitstreams/6096d74e-c8b7-46ce-ad4b-4f02b97f081f/download

Pukthuanthong, K., and Roll, R. (2011). Gold and the dollar (and the euro, pound, and yen). Journal of Banking & Finance, 35(8), 2070–2083.

Rana, H. M. U., and O'Connor, F. (2023). Domestic macroeconomic determinants of precious metals prices in developed and emerging economies: An international analysis of the long and short run. International Review of Financial Analysis, 89, 102813. https://doi.org/10.1016/j.irfa.2023.102813

Reboredo, J. C. (2013). Is gold a safe haven or a hedge for the US dollar? Implications for risk management. Journal of Banking & Finance, 37(8), 2665–2676.

Robinson, Z. (2019). Revisiting gold price behaviour: a structural VAR. Mineral Economics, 32(3), 365–372. https://doi.org/10.1007/s13563-019-00204-4

Sarwar, G. (2012). Intertemporal relations between the market volatility index and stock index returns. Applied Financial Economics, 22(11), 899–909.

Sarwar, G., and Khan, W. (2019). Interrelations of U.S. market fears and emerging markets returns: Global evidence. International Journal of Finance & Economics, 24(1), 527–539.

Shafiee, S., and Topal, E. (2010). An overview of the global gold market and gold price forecasting. Resources Policy, 35(3), 178–189.

Sokhanvar, A., and Hammoudeh, S. (2024). Comparative analysis of responses of risky and safe haven assets to stock market risk before and after the yield curve inversions in the U.S. International Review of Economics & Finance, 94. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4601513

Thaver, R. L., and Lopez, J. (2016). Unemployment as a determinant of gold prices: Empirical evidence. The International Journal of Business and Finance Research, 10(4), 43–52.

Toda, H. Y., and Yamamoto, T. (1995). Statistical inference in vector autoregressions with possibly integrated processes. Journal of Econometrics, 66(1–2), 225–250.

Valadkhani, A., and O'Mahony, B. (2024). Dynamic hedging responses of gold and silver to inflation: A Markov regime-switching VAR analysis. International Review of Economics & Finance, 96, 103741. https://doi.org/10.1016/j.iref.2024.103741

Whaley, R. E. (2000). The investor fear gauge. Journal of Portfolio Management, 26(3), 12–17.

Whaley, R. E. (2009). Understanding the VIX. Journal of Portfolio Management, 35(3), 98–105.

Wu, D. (2025). Institutional adoption and correlation dynamics: Bitcoin's evolving role in financial markets. arXiv preprint arXiv:2501.09911. https://arxiv.org/abs/2501.09911