UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

DO STOCK PRICES BEHAVE LIKE A MARTINGALE, OR IS A FUNDAMENTAL PRESENT-VALUE MODEL A MORE APPROPRIATE DESCRIPTION OF THEIR DYNAMICS AND PREDICTIVE STRUCTURE?

Authors:
Elia Landini, Lorenzo Alessandro Uberti Bona Blotto, Simon Mitrofanoff

1. Introduction

A pivotal question centred around contemporary assets, which has provoked a huge influx of academic literature, is whether assets prices should be modelled as a martingale process, which falls in line with the Eugene Fama Informational Efficiency Hypothesis (1965), or whether they should be anchored to economic fundamental approaches such as the present value relations. If prices follow martingale, then conditional on available information, expected returns must be equated to zero and all predictable components of price are purely based on past current and past market information, this thereby in turn eliminates all potential predictability of future price movements promoting the idea that markets instantaneously incorporate new information. On the other hand, the fundamental approach asserts that prices reflect discounted expectations of future cashflows, there-in through their dividends and subsequently must satisfy the idea that present value identity linking valuation ratios to predictable variation in returns or cash-flow growth. This debate sits at the heart of financial economics because it captures intrinsically different concepts on how markets process information and whether returns can be forecasted. We test these hypothesis empirically by applying both martingale-based tests and the present value modelled detailed in John H Cochrane's (1992) to the Nasdaq-100 Index and singular stock data at hourly, daily, weekly and monthly frequency with the aim to analyse whether markets evolve like a fair game or whether they exhibit systematic deviations consistent with economic fundamentals or behavioural dynamics.

We've divided this project into two distinct parts, where Part 1 will focus on the martingale-based tests of informational efficiency, examining whether the NASDAQ price dynamics as well as constructed synthetised NASDAQ index are consistent with Random Walk (RW1, RW2, RW3) implications taught in the course in order to evaluate whether returns behave like a martingale - difference sequence. The tests which we implement are designed as such in order to assess the weak - and strong for efficiency hypotheses and reveal where real world data deviates from the 'idealised' martingale benchmark. In part 2 we shift our focus onto the present value model developed by Cochrane consumption-based asset pricing framework and cointegration methods to examine whether prices, dividends, and other macroeconomic fundamentals share stable long-run relationships. This section moves beyond the martingale model by testing dynamic restrictions implied by representative-agent models and assessing whether equity valuations anchor to economic fundamentals over the long run.

Part 2 of out project is a development of the already existing theoretical framework resented by (Campbel and Shiller, 1988), we do this through the formularization of dividend growth, discount rates and also subsequently valuation ratios and how they jointly determine assets prices in a dynamic setting. We decided to adopt a reduced form popularised by Cochrane and implement it through a parsimonious VAR in monthly excess return, 12-month cumulated dividend growth, and the dividend price ratio. The idea behind this structure being that it would provide us with a tractable link observable valuation ratios and future cash flows and returns. Hence this section sets out the theoretical foundations that will guide the comparison with the martingale model and the interpretation of mispricing in the subsequent sections.

2. Data Retrieval and Manipulation

We decided to construct our own empirical dataset for this project so that we could effectively cater it to support the two distinctive strands of analysis which we propose, being the martingale tests of informational efficiency, and the Present Value Model and long run fundamental valuation tests. In order to ensure that we can comparability and consistency within our data we collected and processed to a unified scheme.

Data Sources

All equity prices and dividends series for our chosen Nasdaq 100 index was collected using the NASDAQ Data Link which we then independently cross validated using the Yahoo Finance ticker to ensure accuracy in dividend history and also adjusted closing prices. The Nasdaq -100 index series was retrieved directly from Nasdaq Data Link under the ticker NDX. Shares-outstanding data and market-capitalisation figures were sourced from NASDAQ historical fundamentals, Refinitiv, or Yahoo Finance key statistics, depending on availability. The risk-free rate was taken from the 3-month U.S. Treasury Bill yields published by the Federal Reserve Economic Data (FRED) database.

Construction of the Return Data for Martingale Analysis

For the martingale tests (Part 1), the same log returns data is utilised in Part 2, we required high frequency data to evaluate the implications of RW1, RW2 and RW3. Subsequently we proceeded by collecting the Nasdaq-100 index prices and individual stocks prices at hourly, daily, weekly, and monthly frequencies. All raw prices were converted into adjusted closing prices to include the effect to include the effect of dividends and corporate actions, thereby allowing us to focalise not only on the weak EMH but also the semi – strong EMH. These were in turn transformed into natural logarithms:

$$p_{i,t} = \log\left(P_{i,t}\right), \;\; r_{i,t} = p_{i,t} - p_{i,t-1}$$

Hourly data was cleaned to account for the overnight gaps and zero-volume interval, which could have arised while on the other side we cleaned both daily and weekly data for missing values and ensured that it was aligned with the official NASDAQ closing times. Missing observations had to then be manually inspected and for those which constituted large gaps in data, were omitted to in order to ensure that we preserved the serial dependence structure necessary for the martingale testing. The data set that we constructed as a result of this allowed us to form the basis for the variance-ration tests, auto-correlation and martingales difference hypothesis testing for part 1. As a Pre-test we conducted stationary test in order to ensure that or series followed an I(1) process for prices and also that returns followed a white noise process we can seems these non -stationarity in prices by observing the respective time series in figure 1, the white noise process or returns in figure 2 in the appendix and figure 3 with the respective histograms.

Construction of the Dividend consistent Data for the Present Value Model

We account for the fact that the PVM requires a meaningful and strictly positive dividend process, accounting for this we reconstructed the Nasdaq-100 index such that it is consistent with its theoretical assumptions. We removed all Nasdaq 100 firms that did not pay dividends on a regular basis in order to avoid foreseeable mechanical distortions caused by zero dividends, which could have mad ethe price dividend ratio undefined and/or economically meaningless. Let:

$$J_t \equiv C_t^D$$

Denote the set of dividend paying firms. Using only these firms, we constructed a value weighted synthetic index. Let:

$$\left(S_t^i\right)_{t\geq 0}, \quad i \in J_t$$

Denote the adjusted price processes of the dividend paying firms. We then subsequently constructed the deterministic weights as follows:

$$w_{i,t} \equiv \frac{Market\ Cap_{i,t}}{\sum_{j \in J_t} Market\ Cap_{j,t}}, \qquad Market\ Cap_{i,t} \equiv S_t^i \times SharesOutstanding_{i,t}$$

This then allowed us to define the synthetic dividend filtered index as:

$$I_t \equiv \sum_{i \in J_t} w_{i,t} S_t^i$$

This framework setup which was inline with that of the log linearised setup proposed by Cochrane in 1992 ensured that the index reflects only the firms which had economically meaningful dividend streams.

Lastly, since dividends for the stocks within the Nasdaq index are not normally paid monthly and many firms distribute quarterly, we constructed a trailing 12-month dividend measure for each firm. The trailing-dividend measure $D_t^{(12)}$captures the level of dividends over a longer horizon than the raw monthly flow and provides a more stable proxy for the long-run dividend process. Under this transformation, cointegration between $d_t$and $p_t$can be interpreted as a long-run equilibrium relationship between the log price and log dividend processes, while simultaneously addressing the mechanical relationship between price and dividends and the seasonality of dividend payments. Let:

$$(D_t^i)_{t \geq 0}$$

Denote the dividend process of firm i. And then defining the trailing dividend as:

$$D_t^{i,(12)} \equiv \sum_{k=0}^{11} D_{t-k}^i, \qquad d_t^{i,(12)} \equiv \log D_t^{i,(12)}$$

The index level trailing dividend uses the same deterministic weights:

$$D_t^{(12)} \equiv \sum_{i \in J_t} w_{i,t} D_t^{i,(12)}, \quad d_t^{(12)} \equiv \log D_t^{(12)}, \quad \Delta d_{t+12}^{(12)} \equiv d_{t+12}^{(12)} - d_t^{(12)}$$

The log price of the synthetic index is:

$$p_t \equiv \log I_t$$

And hence the valuation ratio used in the PVM analysis is:

$$pd_t \equiv \log\left(\frac{D_t^{(12)}}{P_t}\right) = d_t^{(12)} - p_t$$

To test for cointegration, we implement the Engle–Granger two-step procedure. First, we estimate the long-run regression formulated as follows:

$$p_t = \alpha + \beta d_t + u_t,$$

and then apply an augmented Dickey-Fuller (ADF) test to the residuals u_t. As an additional robustness check, we also compute an ADF test directly on the spread $d_t - p_t$, following the practice in Campbell (1988). Finally, to corroborate the Engle–Granger results, we employ the Johansen cointegration test on the bivariate system (d_t, p_t).

Final Dataset for Empirical Application

After cleaning, filtering ad lastly aligning our series: we had our data set which was applicable to the Martingale analysis which uses the index and stock level log returns at multiple frequencies to test the RW1 to RW3, serial independence, and martingale difference implications. And also, a data set for the PVM model which uses monthly log prices , trailing log dividends and with the constructed valuation ratios to test the: Engle-Granger Cointegration, Johansen cointegration, pd_t stationarity and finally VAR decompositions of expected returns and dividend growth (Cochrane, 1992). Descriptive Statistics are found in Table 1 and 2 of the Appendix. Furthermore, the evolution of descriptive statistics is captured from Figure 1 to Figure 8.

3. Theoretical Framework

3.1 Part 1 - Theoretical Framework

In this section we being to outline the theoretical framework of the efficient market hypothesis (Fama 1970), the martingale hypothesis with details on super/submartingales and the three Random walk hypotheses.

Efficient Market Hypothesis (EMH)

A capital market is said to be an efficient one if and only if it fully and correctly reflects all relevant information in determining prices. Formally, the market is said to be efficient, with respect to some information set $\mathcal{F}_t$ where $\mathcal{F}_t = \{ P_t, P_{t-1}, P_{t-2}, \dots, P_0\}$, if prices ($P_t$) are unaffected by the revelation of new information. Thereby, efficiency with respect to an information set, $\mathcal{F}_t$, implies that it is impossible to make economic profits by trading on the basis of $\mathcal{F}_t$. There are the distinctions between weak, semi-strong and strong form efficient markets illustrated by Roberts (1967). The weak form of the EMH implies that prices fully reflect the information contained in the historical sequence of prices and therefore as mentioned before investors cannot devise an investment strategy to yield abnormal profits on the basis of an analysis of past price patterns, this is also commonly referred to the Random walk hypothesis. The semi-strong EMH leans on the idea that current stock prices not only reflect the past information by also publicly available information relevant to companies' securities, therefore by analysing company balance sheets, income statements, etc, it will not lead to abnormal profits. And finally, the strong EMH asserts that all information that is known to any economic agent is seen in the prices. We focus specifically on the weak and semi strong EMH in this project.

Martingale

The first formal economic argument of the EMH was formalized by Samuelson as he moved away from Fama (1965) idea to model with random walk and focussed on the martingale property in his paper. Let $\{P_t\}$ be a stochastic price process with the information set $\mathcal{F}_{t-1}$ containing all past values of the process up till time t-1. P_t is said to be a martingale if the following properties are satisfied:

$$\mathrm{E}_t[\,|P_t|] < \infty, \forall t$$

$$\mathrm{E}_t[P_t \mid \mathcal{F}_{t-1}\,] = P_{t-1}\,, \forall t$$

Whereby the increments form a fair game: $\Delta P_t = P_t - P_{t-1}$ *then* $\mathrm{E}_t[\Delta P_t \mid \mathcal{F}_{t-1}\,] = 0$. This concept allows us to define the main testable element which states that if the stochastic price process is martingale, then the return for all t will be equal to 0. [proof 1]. Super and Sub martingales are similar variants to the main martingales but they state that:

$$\mathrm{E}_t[P_t \mid \mathcal{F}_{t-1}\,] \le P_{t-1}\ (supermartingale) \land\ \mathrm{E}_t[P_t \mid \mathcal{F}_{t-1}\,] \ge P_{t-1}\ (submartingale)$$

The importance of these being that they generalize the martingale (equality) to allow for systematic drifts. Within our project we outline aswell a subsection of the Nasdaq as mentioned before where we filter out non-dividend paying stocks and only retain the dividend paying ones. The martigale property of this Synthesised Index is still maintained as detailed in proof 2.

Random Walk – RW1, RW2, RW3

Let $\{P_t\}$ be the price process and define the increments $\varepsilon_t = P_t - P_{t-1}$. A random walk is a process characterised as:

$$P_t = P_{t-1} + \varepsilon_t \, , for \; t = 1,2, \dots \rightarrow \; with \; restcitions \; on \; the \; sequence \; \{\varepsilon_t\}$$

Fama (1970) draws the distinction:

- Random Walk 1 (RW1), $\{\varepsilon_t\}$ are i.i.d with finite variance:

$$\{\varepsilon_t\} \; i.i.d, \quad \mathrm{E}_t[\varepsilon_t] = 0 \, , \; Var[\, \varepsilon_t \,] \; = \; \sigma^2 < \; \infty$$

- Random Walk 2 (RW2), $\{\varepsilon_t\}$ are independent (but not necessarily identically distributed):

$$\varepsilon_t \; \perp \; \varepsilon_s \; for \; t \; \neq s \, , \qquad \{\varepsilon_t\} \; i.i.d, \quad \mathrm{E}_t[\varepsilon_t] = 0 \, , \; Var[\, \varepsilon_t \,] \; = \; \sigma_t^2$$

- Random Walk 3, $\{\varepsilon_t\}$ form a martingale difference sequence:

$$\mathrm{E}\,[\, \varepsilon_t \mid \mathcal{F}_{t-1} \,] = 0, \qquad Var(\varepsilon_t) = \; \sigma_t^2$$

However, it is important to note that logically the martingales property can only be applied to the risk neural world, in the much more realistic risk averse world agents require a risk premium in order to account for risk and hence the martingale falls through.

Methodology for Testing the Martingale and Random Walk Hypotheses and Results

- ***Mean Return t-test (Drift Test)***

The first step in our evaluation for the martingale hypothesis is to examine whether the unconditional mean return of the asset is statistically different from zero. A martingale requires that the conditional expected price change is zero thereby implying that expected return, when controlling for the equilibrium risk premium, should not exhibit a systematic drift. Since we used a one-sample t-test on the sample mean return we can thereby evaluate the following 3 hypothesises: whether returns are significantly positive (submartingale behaviour), significantly negative (supermartingale behaviour) or indistinguishable from zero (martingale consistent) . This allows us to directly assesses whether assets prices follow a "fair game" property, making it a foundational empirical check for the weak EMH. The results of the Drift test can be found in Table 3 in the Appendix.

The results depict that there a clear dependence of the validity of the martingale hypothesis on the sampling frequency. At the hourly horizon, the NASDAQ log-returns show no statistically significant mean drift: the estimated mean return is extremely small, the t-statistic (0.801) is far from conventional critical values, and both one-sided tests fail to reject the null of zero drift. This supports martingale-compatible behaviour at high frequencies, consistent with the idea that, in liquid markets with continuous trading, short-term returns are unpredictable and price changes resemble a fair game. However, as the horizon lengthens, the results change markedly. For daily, weekly, and monthly returns, the estimated mean return becomes economically and statistically significant. Across all three frequencies, the t-statistics are sizeable (≈2.4–2.9 for NASDAQ; >3 for the synthetic series), and the one-sided p-values overwhelmingly reject the $H_0: \mu = 0$ in favour of $\mu > 0$. This implies a persistent positive drift, meaning that log prices follow a submartingale rather than a true martingale. In general, these findings indicate that the stringent martingale hypothesis is overly rigorous for real-world asset pricing. Although it effectively captures short-term return dynamics, positive drift builds up over time, leading prices to behave like a submartingale instead of a martingale. This falls in line with standard assets pricing theory whereby assets require a risk premium promoting further the know result that martingales are only relevant in the risk neutral world.

- ***Cowles – Jones Test (1937)***

The Cowles – Jones test evaluates whether the sequence of positive and negative returns is independent over time by comparing the frequency of “sequences” (same sign) and “reversals” (opposite sign).

$$\widehat{CJ} \equiv \frac{N_s}{N_r}$$

Under the previously stated assumptions of RW1, returns must be independently distributed which therefore trivially that the sequences and the reversals should occur with equal probability. If deviations from this benchmark are observable then it implies that there could be the presence of short- term predictability in the direction of price movements, violating both the independence assumption of RW1 and the fair game property of the martingale hypothesis. Allowing to see if early evidence suggests price changes behave like i.i.d innovation processes, as required for the strongest form of RW and for the strictest version of the weak EMH. The results of the Cowles – Jones Test can be found in Table 4 of the Appendix.

As previously mentioned, the Cowles – Jones statistic provide evidence on whether price changes exhibit the sign-independence (a requirement for RW1), i.e. i.i.d. increments, and logically as a results whether returns behave like the innovation process of a strict martingale. Across all frequencies for the NASDAQ, the runs test consistently shows that the proportion of same-sign sequences relative to reversals does not differ significantly from the benchmark value of 0.5. The z-statistics are small in magnitude and the associated p-values are well above standard significance levels. Therefore, for the NASDAQ, we cannot reject RW1 at hourly, daily, weekly, or monthly horizons. This implies in terms of a directional change, the NASDAQ shows independent signs of return, therefore no detectable short-term predictability. For the synthetic Nasdaq series, which was constructed as a positively trending random walk with independent increments, the results behave as expected: the runs test fails to reject RW1 at most frequencies but does reject RW1 at the weekly and monthly horizons. These rejections reflect the fact that a strong positive drift increases the likelihood of consecutive positive returns, thereby generating longer same-sign sequences. The test therefore correctly identifies the inherent directional persistence induced by the imposed drift.

- ***Drift t-Test at Multiple Horizons (Campbell Log Horizon Test)***

Although returns from a single period may not show any noticeable drift, deviations from the martingale property can occur when returns are combined over extended periods. The Campbell long-horizon t-test creates multi-period returns (e.g., h = 5, 20, 60) and assesses if their averages deviate from zero. According to the martingale hypothesis, the overall expected return across h periods should stay at zero, indicating that price shifts over long horizons must remain unforecastable. The emergence of a statistically significant drift solely at extended horizons indicates the existence of lasting trends or average growth in log-prices that challenge the martingale model yet go unnoticed in short-term data. Consequently, these tests enhance the one-period findings by investigating if efficiency violations occur after short-term fluctuations are averaged out. These results are found in table 5 of the appendix.

The long-horizon drift results offer a more distinct understanding of how departures from martingale behaviour arise when returns are combined over several periods. At the hourly level, none of the multi-period return intervals (h = 1, 5, 20, 60) show noteworthy drift. All t-statistics stay low, and their associated p-values significantly surpass standard significance thresholds. This supports the finding from the one-period drift tests: at extremely high frequencies, price fluctuations act in accordance with the martingale hypothesis, showing no observable predictable element in average returns. For daily returns of NASDAQ, a distinct pattern emerges. Although the horizons h = 1 and h = 5 do not reject the martingale null hypothesis, the drift reaches statistical significance at h = 20 and h = 60. This suggests that while short-term returns may appear to be random, consistent upward trends appear when price changes are combined over several weeks or months. The behaviour is thus inconsistent with strict martingale dynamics, where expected multi-period returns should stay zero across all time horizons. Rather, the evidence indicates a submartingale behaviour, influenced by a minor yet consistent upward trend that accumulates over time. The NASDAQ outcomes for the week and month reinforces this assertion. By h = 5, the long-horizon drift is noticeable, with very strong rejections at h = 20, and in nearly all instances, also at h = 60. These results indicate that long-term returns experience a consistent positive trend, despite the short-term changes seeming random. Results can be found in table 6 of the appendix. .

- ***ADF Unit Root Tests on Log Prices***

The Augmented Dickey–Fuller (ADF) test is used to assess if the log-price series has a unit root, which characterizes a Random Walk II (RW2). A martingale in prices indicates that log prices are non-stationary, changing over time with innovations that build up without mean reversion. Not rejecting the unit root hypothesis aligns with the martingale framework. Conducting ADF tests both with and without deterministic components (constant, linear trend) enables us to differentiate a pure random walk from a random walk with drift and a trend-stationary process. If the series is stationary, it would completely dismiss the martingale model, suggesting a predictable long-term behaviour that contradicts market efficiency.

Across all sampling frequencies, the ADF statistics on NASDAQ and synthetic log-prices almost uniformly fail to reject the unit–root null. For the NASDAQ series, the ADF with only a constant yields very high p-values at the hourly, daily, weekly and monthly horizons ($p \approx 0.88$–0.99), so the data are fully consistent with a RW2 specification in which log prices follow a random walk with drift. When a deterministic trend is added, the test statistics become more negative but remain above the 5% critical values (p between about 0.06 and 0.57), again implying that we cannot reject a stochastic trend in favour of trend-stationarity. The synthetic log-price series behaves similarly: with a constant, the ADF statistic is clearly non-rejection at all horizons; with a trend, the test moves closer to the rejection region, but the p-values (around 5–10%) still do not cross conventional significance thresholds. Overall, the ADF evidence points to non-stationary log prices with a unit root for both NASDAQ and the synthetic benchmark. This is broadly compatible with the martingale / RW2 view of asset prices, and there is no strong indication, on the basis of standard ADF critical values, that prices are stationary around a deterministic trend. Results are found in table 7 through to 10.

- ***DF Regression $\Delta P_t = \alpha + \phi P_{t-1} + \varepsilon_t$***

The DF regression offers a parametric assessment of whether the autoregressive coefficient φ on previous prices equals zero, as suggested by the random walk model. According to the martingale hypothesis, price variations should be independent of prior price levels, and the coefficient φ must not be statistically negative. This regression also directly estimates α and β (in the trend-augmented specification), enabling us to identify deterministic drift or trend elements that do not align with a martingale. The DF regression thus offers a more structural method than the typical ADF test, making it essential for checking if price fluctuations act like a martingale difference sequence originating from a unit-root process. The DF regressions offer a more structural look at the same question by directly estimating the autoregressive coefficient on lagged prices. In the specification without trend, ϕ is always very close to zero and statistically insignificant for both NASDAQ and synthetic prices at all horizons (t-statistics in absolute value < 1), so we again fail to reject the unit-root restriction implied by RW2. However, once a linear trend is included, the picture changes: the trend coefficients β are positive and highly significant, and the estimated ϕ becomes negative and statistically different from zero ($t \approx -2.1$ for hourly NASDAQ and around –3.1 to –3.7 for the daily, weekly and monthly series, as well as for all synthetic horizons). Interpreted literally, these results point towards trend-stationarity: after removing a deterministic upward trend, log prices display mean reversion rather than a pure unit root. This suggests that part of price dynamics can be captured by a deterministic growth path (reflecting, e.g., long-run expected returns), around which deviations are stationary. From a martingale perspective, this weakens the case for a strict RW2 in raw log prices and instead hints that any martingale-difference behaviour may hold only for appropriately detrended price series or for returns rather than levels. These results are also shown in table 7 through to 10.

- ***Autocorrelation Function (ACF) and Ljung – Box Q – test***

According to Random Walk III (RW3) and the martingale difference hypothesis, asset returns should show no linear correlation at any lag. The sample autocorrelation function (ACF) examines correlation at specific lags, whereas the Ljung–Box Q-statistic assesses whether returns are collectively uncorrelated over several lags. Identifying notable autocorrelation suggests that some of today’s returns can be forecasted using previous returns, contradicting the martingale difference property and offering clear evidence against weak-form EMH. Consequently, these tests focus on the fundamental prediction of market efficiency: that historical price changes hold no actionable insights regarding future returns.

Across all sampling frequencies, the autocorrelation and Ljung–Box tests reveal that return dynamics deviate from the martingale-difference predictions of RW3 in several important ways. At the hourly and daily frequencies, NASDAQ returns display statistically significant joint autocorrelation, with Ljung–Box p-values well below 1%, indicating that recent returns help predict current returns at short horizons. Individual ACF coefficients are small but not collectively zero, confirming a violation of the martingale difference property. The synthetic series behaves similarly: although it was designed as a drifting random walk, its returns show detectable autocorrelation at daily and weekly horizons, again contradicting RW3. However, the behaviour changes at lower frequencies. For both NASDAQ and the synthetic benchmark, weekly and monthly returns show no statistically significant autocorrelation under Ljung–Box Q-tests (p-values > 0.25), suggesting that once high-frequency noise is aggregated, return innovations appear closer to white noise. Overall, the autocorrelation evidence shows that RW3 fails for high-frequency returns but becomes more plausible at weekly and monthly horizons, where return dynamics are largely free from exploitable serial dependence.

- ***Variance Ratio Tests (Lo – Mackinlay)***

The variance ratio (VR) test is among the most effective methods for differentiating between random walk patterns and mean reversion/momentum influences. In a martingale or RW3 framework, the variance of returns over q periods must be precisely q times the variance of returns over one period. Variations from this scaling characteristic suggest either positive serial correlation (momentum: VR > 1) or negative serial correlation (mean reversion: VR < 1). These patterns are in conflict with the unpredictable-growth characteristic of martingales. The VR test is thus crucial for evaluating if price fluctuations adhere to the diffusion process suggested by EMH or if returns display a consistent pattern over different timeframe.

The variance ratio results reinforce the conclusion that market efficiency and RW3 validity depend strongly on the sampling frequency. For NASDAQ hourly returns, VR statistics at short horizons (q = 2, 5) are consistent with RW3, but at longer horizons (q = 10) the VR drops below 1 and becomes highly significant, indicating mean reversion in intraday price movements. Daily returns offer the strongest rejection of RW3: VR statistics are far below 1 at all horizons tested, with extremely large negative z-scores, signalling pervasive negative autocorrelation and strong mean-reverting behaviour. Weekly returns also reject RW3 at all VR horizons for both NASDAQ and synthetic returns, again suggesting persistent mean reversion. In contrast, monthly returns for both datasets show VR statistics insignificantly different from 1 across all horizons, implying that long-horizon return variance scales proportionately with time—consistent with the diffusion behaviour predicted by RW3. Taken together, the VR tests imply that while daily and weekly returns exhibit systematic departures from martingale dynamics, long-horizon monthly returns behave in a way that is statistically indistinguishable from an efficient martingale-difference process.

- ***Summary***

Collectively, the comprehensive set of martingale and random-walk tests provides a detailed view of how closely NASDAQ price movements correspond with the weak-form Efficient Market Hypothesis (EMH). At extremely high frequencies, especially the hourly level, returns display characteristics typical of a martingale difference sequence: mean drift is statistically similar to zero, sign changes seem independent, long-horizon drift tests reveal no anomalies, and ADF/DF models without trend strongly endorse a unit-root representation of log prices. This suggests that at high frequencies, price movements appear to be fair-game innovations lacking any systematic predictability. Nonetheless, as the length of the sampling interval increases, significant deviations from pure martingale behaviour appear. The drift tests for daily, weekly, and monthly returns consistently reveal a lasting positive drift, indicating that log prices behave like a submartingale instead of a genuine martingale. Extended drift tests reinforce this conclusion, indicating that although short-term movements are hard to predict, total returns over several weeks or months reveal statistically significant patterns. Even though ADF tests persistently indicate a unit root in log prices, the DF regressions featuring deterministic trends demonstrate consistent mean reversion after a trend component is removed, implying that some of the seen price behaviour represents a deterministic long-term growth trajectory rather than solely stochastic diffusion. Serial dependence tests also suggest frequency-related variations from the martingale difference hypothesis. At both hourly and daily intervals, notable autocorrelation and Ljung–Box test results indicate potential short-term predictability, while weekly and monthly returns show minimal correlation, aligning better with EMH expectations. The variance–ratio tests validate this trend: robust mean reversion at daily and weekly intervals contrasts with monthly returns, where the variance increases in proportion to time and aligns with a random-walk diffusion model.

In general, the data suggests a distinct conclusion: the martingale and random-walk theories adequately represent high-frequency price movements, but their accuracy diminishes as the time frame extends. Positive drift, weak serial correlation, and mean reversion accumulate over extended periods, leading to consistent divergences from fair-game behaviour. These findings align with conventional asset-pricing theory, which suggests that risk premia lead to upward trends, as well as with empirical studies that reveal short-term microstructure impacts and medium-term return reversals. Thus, although the strict martingale model serves as a valuable theoretical reference point, actual NASDAQ prices display more complex dynamics that are fundamentally influenced by the sampling frequency.

3.2 Part 2 Theoretical Framework Layout – Fundamental Value Model

<u>Present-Value Identity and Reduced-Form Implementation</u>

In the log-linear framework of Campbell and Shiller (1988), the dividend–price ratio satisfies the present-value identity:

$$dp_t = \kappa + \mathbb{E}_t[\sum_{j=1}^{\infty} \rho^{j-1}(\Delta d_{t+j}^{(12)} - r_{t+j})],$$

where $\kappa$is a constant and $0 < \rho < 1$is a log-linearization parameter determined by the average dividend-price ratio. This identity states that a high dividend-price ratio must reflect either low expected future dividend growth or high expected future discount rates (returns), or both. Rather than explicitly evaluating the infinite sum on the right-hand side, we will adopt a reduced-form implementation: a VAR is estimated for excess returns, log 12-month-cumulated dividend growth, and dividend-price ratio. The VAR solution implicitly encodes the joint dynamics of expected future dividends and returns that underlie the present-value relation. In this sense, the VAR-implied path of the dividend-price ratio provides a model-consistent measure of "fundamental value" in the spirit of the Campbell-Shiller-Cochrane framework (2008).

<u>VAR Specification on 12-Month Aggregates</u>

The state vector z_t, at the base of our VAR model, can be formulated as follows.

$$\mathbf{z}_t \equiv \begin{bmatrix} rx_{t+1} \\ \Delta d_{t+1}^{(12)} \\ dp_t \end{bmatrix}$$

where:

$$rx_{t+1} = (p_{t+1} - p_t) + \log\left(1 + \frac{D_t^{(12)}}{P_t}\right) - r_{t+1}^f$$

$$\Delta d_{t+1}^{(12)} = d_{t+1}^{(12)} - d_t^{(12)}$$

$$pd_t \equiv \log\left(\frac{D_t^{(12)}}{P_t}\right) = d_t^{(12)} - p_t$$

To evaluate the forecasting performance of the VAR model, the monthly dataset is partitioned into a training and a testing subsample. All observations up to December 2021 are assigned to the training set, which is used to estimate the VAR parameters and generate in-sample fitted values. While observations from January 2022 onward constitute the evaluation sample, over which the model is iterated forward to produce unbiased out-of-sample forecasts of the dividend-price ratio.

To determine the appropriate dynamic structure of the model, we conduct a formal lag-order selection exercise using the training sample. Specifically, we estimate a sequence of VAR models with up to ten lags and compute the standard information criteria, AIC, BIC, FPE, and HQIC, for each specification. The results, summarized in Table 21, show that the Akaike Information Criterion (AIC) and the Final Prediction Error (FPE) both reach their minimum at lag 2, indicating that a VAR(2) captures the underlying joint dynamics most effectively. Although the Bayesian Information Criterion (BIC), which strongly penalizes model complexity, points toward a more parsimonious VAR(1), our objective is to optimize out-of-sample predictive performance rather than impose maximal parsimony. For this reason, we prioritize AIC and FPE in selecting the preferred lag length. The idea behind is that with this approach we want to ensure that the model possesses enough flexibility to capture the medium-run interactions among long-horizon returns, dividend growth, and the dividend-price ratio, while still being disciplined by formal selection criteria.

$$\mathrm{z}_t = \Gamma \mathrm{z}_{t-1} + \Gamma \mathrm{z}_{t-2} + \mathrm{u}_t,$$

where $\Gamma_1$and Γ_2 are 3×3coefficient matrices and u_t is a vector of mean-zero innovations with covariance matrix Σ_u. While the present-value literature often relies on parsimonious VAR(1) structures for analytical tractability (e.g., Golez et al., 2015), our lag-selection results, just shown indicate that a VAR(2) provides superior in-sample fit and improves out-of-sample forecasting performance. We therefore adopt this specification to better capture the medium-run dynamics. The VAR(2) is estimated on the training sample of synthetic index data, and for each month the model produces an in-sample fitted value $\widehat{dp}_t$of the dividend-price ratio. Over the evaluation period, the estimated VAR is iterated forward one step at a time to generate out-of-sample forecasts of dp_t, which we also denote by $\widehat{dp}_t$for notational simplicity.

Present-Value Benchmark Price and Mispricing

Using the fitted outputs of the VAR, we reconstruct an implied path for dividends and prices, fully consistent with the model's present-value structure. First, the fitted log dividend-growth forecasts, $\widehat{\Delta d}_{t+1}$, are exponentiated to obtain the gross growth rates,

$$g_{t+1} = e^{\widehat{\Delta d}_{t+1}}.$$

Then, starting from the known initial dividend level at time *t=0*, $D_0 = 0.0336015$, we generate the model-implied dividend sequence recursively according to

$$D_t^{PVM} = D_{t-1}^{PVM} \cdot g_t,$$

which mirrors the log-growth identity $d_t = d_{t-1} + \Delta d_t$, being aware of the evolution of the dividend growth rate over time, which prevents us from using a constant discounted growth rate. Given these reconstructed dividends and the fitted dividend-price ratio $\widehat{dp}_t = \log{(D_t/P_t)}$, we recover the implied price series using:

$$P_t^{PVM} = \frac{D_t^{PVM}}{e^{\widehat{dp}_t}}.$$

By construction, $P_t^{PV}$is the price that would be consistent with the VAR-implied joint expectations of 12-months cumulated dividends and discount rates under the log-linear present-value relation. The mispricing of the synthetic market relative to fundamentals is then defined as:

$$\text{mispricing}_t \equiv P_t - P_t^{PV},$$

$$\text{mispricing\_ratio}_t \equiv \frac{P_t}{P_t^{PV}} - 1.$$

These measures summarize the extent to which the realized synthetic price deviates from the present-value benchmark implied by the VAR. In the remainder of the paper, this fundamental benchmark is compared to a Fama-style martingale benchmark, under which price changes are, by construction, unpredictable given past information.

While the VAR could in principle be used to decompose the variance of dp_t into components attributable to expected-return and expected-dividend-growth news, in the spirit of Cochrane (1992), the present paper focuses solely on the construction of $P_t^{PV}$and the resulting mispricing measures, which are sufficient for the comparison with the martingale model.

Monthly dividends and prices

Now, going back for a moment to monthly price and dividends (instead of 12-month cumulated dividends), Table 13 reports OLS estimates from the regression of the log index price on the log of monthly dividends,

$$p_t = \alpha + \beta d_t + u_t,$$

where $p_t = \log\ P_t$and $d_t = \log\ D_t$. The slope coefficient on $d_t$is positive and highly significant ($\hat{\beta} = 0.5533$, t-statistic = 16.10, $p < 0.001$), and the regression explains about half of the cross-time variation in prices (R-squared = 0.486, adjusted R-squared = 0.484). The F-statistic is large (259.1, $p \approx 1.8 \times 10^{-41}$), indicating that the overall regression is highly statistically significant. At the same time, the Durbin–Watson statistic is 1.289, well below 2, which points to substantial positive autocorrelation in the residuals, as expected when regressing nonstationary series on one another.

To examine whether this relationship reflects a genuine long-run equilibrium, Table 14 reports an Engle-Granger residual-based Augmented Dickey-Fuller (ADF) test applied to $\hat{u}_t$. The ADF statistic is 0.487 with a p-value of 0.984, far above conventional significance levels. The test statistic is well above the 10% critical value of –2.573, so we fail to reject the null of a unit root in the residuals. This implies that, despite the high R-squared, the residuals from the price-dividend regression remain nonstationary and there is no evidence of cointegration between monthly log prices and monthly log dividends. Table 15 applies the "Campbell approach" by testing directly whether the log dividend–price ratio:

$$dp_t = \log\ D_t - \log\ P_t = d_t - p_t$$

is stationary. The ADF statistic of –1.793 (p-value = 0.384) again lies above the 10% critical value of –2.573, so the unit-root null cannot be rejected. This result is consistent with the Engle–Granger test: the monthly dividend–price ratio appears to behave as an $I(1)$process in this sample, suggesting that monthly dividends and prices do not form a stable cointegrating pair.

Trailing 12-month dividends and prices

From the minimal linear relationship between monthly prices and dividends, we now add a layer of complexity by adding the already discussed trailing 12-months dividends to our OLS analysis. Table 16 reports OLS estimates for the regression

$$d_t^{(12)} = \alpha + \beta p_t + \varepsilon_t,$$

where $d_t^{(12)} = \log\ D_t^{(12)}$. The slope coefficient is again positive and highly significant ($\hat{\beta} = 1.1168$, t-statistic = 36.97, $p < 0.001$), and the fit is much tighter: the R-squared is 0.833 (adjusted R-squared = 0.832), and the F-statistic is extremely large (1367, $p \approx 1.7 \times 10^{-108}$). This regression suggests a very strong long-run association between the smoothed dividend measure and prices.

However, the Durbin-Watson statistic of 0.163 indicates extreme positive autocorrelation in the residuals, which raises the possibility that the high R-squared partly reflects a spurious regression between integrated series. This concern is confirmed by the residual-based Engle-Granger test in Table 17. The ADF statistic on the regression residuals is -2.402 with a p-value of 0.141, which remains above the 10% critical value (-2.573). Thus, even with 12-month trailing dividends, we still **fail to reject** the null that the residuals are non-stationary, and therefore do not find robust evidence of cointegration between $p_t$and $d_t^{(12)}$at conventional significance levels. Table 18 reports an ADF test directly on the 12-month dividend–price ratio:

$$dp_t^{(12)} = \log\ D_t^{(12)} - \log\ P_t,$$

which yields an ADF statistic of -1.575 with a p-value of 0.496. This is, again, far from the critical values, and the unit-root null cannot be rejected. Thus, in this sample, the smoothed dividend–price ratio also appears to be nonstationary.

Johansen test and cointegrating vector estimates

To complement the Engle-Granger tests, Table 19 presents Johansen trace and maximum eigenvalue tests for the bivariate system $(p_t, d_t^{(12)})$. The trace statistic for the null of no cointegration ($r = 0$) is 15.268, which falls marginally below the 5% critical value of 15.494, implying no cointegrating vector at the 5% level according to the trace test (and likewise at the 1% level). By contrast, the maximum eigenvalue statistic for $r = 0$is 15.158, which slightly exceeds its 5% critical value of 14.264, suggesting the existence of a single cointegrating relationship at the 5% level in the max-eigen test. For $r \leq 1$, both trace and max-eigen statistics are far below the corresponding critical values, consistent with at most one cointegrating relation.

Taken at face value, the Johansen tests deliver mixed and borderline evidence: the max-eigenvalue test points to one cointegrating vector at the 5% level, whereas the trace test falls just short of significance. Table 20 reports the estimated eigenvector β, with raw estimate $[1.7163, -2.4949]$and normalized representation $[1, -1.4536]$when imposing $p_t = 1$. This corresponds to a putative long-run relation of the form

$$p_t - 1.4536\ d_t^{(12)} = \text{constant} + \text{stationary error.}$$

Given the conflicting Johansen test statistics and the failure of the Engle–Granger and direct ADF tests to reject nonstationary of the dividend–price ratio, this cointegrating vector should be interpreted with caution and regarded, at best, as weak evidence of a long-run equilibrium relation between prices and 12-month dividends.

Implications for the subsequent analysis

Taken together, the pre-tests suggest that, in this synthetic Nasdaq-based dataset, log prices and both raw monthly dividends and 12-month trailing dividends are best viewed as highly persistent, possibly $I(1)$processes without strong, robust evidence of cointegration at standard significance levels. The dividend-price ratio, whether constructed with monthly dividends $dp_t = \log\ D_t - \log\ P_t$ or with 12-month dividends $dp_t^{(12)} = \log\ D_t^{(12)} - \log\ P_t$, behaves empirically as a near-unit-root process rather than a clearly stationary spread, but the second avoids the typical distribution of series affected by periodical seasonality, which is the case for quarterly issued dividends.

This finding contrasts with the conventional assumption in much of the U.S. aggregate data literature that dividends and prices are cointegrated and that the dividend-price ratio is stationary (e.g. Campbell and Shiller, 1988; Cochrane, 1992), but is not unprecedented in shorter samples, sector-specific data, or synthetic settings. It motivates the subsequent use of a VAR in levels for $\left(rx_{t+1}, \Delta d_{t+1}^{(12)}, dp_t\right)$, rather than imposing a cointegrating structure a priori. In particular, we treat the dividend-price ratio as a highly persistent predictor consistent with the log-linear present-value framework, while refraining from assuming that it is strictly $I(0)$ in this sample. This choice preserves comparability with the Campbell-Shiller-Cochrane present-value VAR approach, but we also acknowledges the empirical limitations of the available data.

4. Results

VAR order selection and dynamics

The VAR(2) estimated on the training sample provides a comprehensive view of the joint dynamics of long-horizon returns (rx_{t+1}), dividend growth (Δd_{t+1}), and the dividend-price ratio (dp_t). The information criteria strongly support the inclusion of two lags (AIC, FPE, HQIC), with only BIC favoring a more parsimonious VAR(1). Given our focus on maximizing forecasting performance, the VAR(2) specification is retained. For notation simplicity the earlier defined trailing 12-months dividend will be recalled as just d_t, together with its derived-metrics.

Starting from the return equation, we clearly note strong autoregressive behavior, as displayed in Table 22. The first lag of returns has a coefficient of 0.6484 with a t-statistic of 10.053, while the second lag contributes an additional 0.3160 with a t-statistic of 4.892. Together, these coefficients imply a total persistence of roughly 0.964, indicating that long-horizon returns in the synthetic index are highly persistent from month to month. By contrast, dividend growth and the dividend-price ratio offer little to no predictive power for returns. Their coefficients across both lags are economically small, ranging from -0.1176 to 0.2961, and statistically insignificant, with absolute t-statistics all below 0.6 and p-values well above 0.5. The constant term, -0.1313 (t = -0.356), is also insignificant. These results indicate that, within this dataset, return predictability arises almost entirely from autoregressive dynamics rather than from cash-flow information or valuation ratios.

On the other hand, the dividend-growth equation shows substantially weaker structure (Table 23). The constant term is -0.3177 with a statistically significant t-statistic of -2.677, indicating a negative drift in the log 12-months cumulated dividend-growth process. Beyond this drift component, the dynamics are largely dominated by noise: the first lag of 12-months cumulated dividend growth is -0.1077 (t = -1.604), and the second lag is 0.1770 (t = 0.676), neither reaching conventional significance levels. Returns and the dividend-price ratio contribute minimally, with coefficients near zero (e.g., 0.00238 for rx_{t+1}at lag 1 and -0.2293 for dp_tat lag 1) and t-statistics well under 1. Again, model results are consistent with the established difficulty of predicting dividend growth, which typically behaves as a weakly autocorrelated and highly noisy process in both real and synthetic index data, embedding its intrinsic seasonality.

The most informative dynamics arise in the dividend-price ratio equation (Table 24). The first lag of Δd_{t+1}dominates the equation, with a coefficient of 0.9962 and an exceptionally high t-statistic of 57.687, reflecting the mechanical relationship implied by the identity $dp_t = d_t - p_t$. The dividend-price ratio itself is also highly persistent: the first lag enters with a coefficient of 1.0375 and a t-statistic of 15.880, confirming slow-moving valuation ratios. Interestingly, both lags of returns play a statistically significant but offsetting role. The first lag of returns enters positively (0.0139, t = 2.601), suggesting that recent high returns temporarily lift the dividend–price ratio, while the second lag enters negatively (–0.0165, t = –3.083), indicating a subsequent correction. These dynamics reflect short-run adjustment behaviour consistent with price movements feeding through the valuation ratio with delays.

Last, the residual correlation matrix shows generally low cross-equation correlations, suggesting a rather good specification of the model. The only moderate correlation is between returns and the dividend-price ratio (0.197), which is expected given the mechanical link between price moves and valuation ratios. Dividend-growth residuals are nearly uncorrelated with both returns (-0.0137) and dp_t(-0.0304), reflecting the noisy nature of dividend-growth dynamics and the absence of omitted shared drivers. Taken together, our VAR(2) model seems to deliver internally coherent dynamics: returns exhibit strong persistence, dividend growth remains difficult to predict, and the dividend-price ratio reacts systematically to both its own past values and to the components of the present-value identity.

VAR impulse response functions

From the VAR(2) framework, we have then computed the relative impulse response functions (Figure 10). Our analysis reveals that, a return shock generates a large and immediate positive response in rx_{t+1}, consistent with the strong autoregressive structure estimated in the VAR, and the effect decays gradually over roughly a year. Neither trailing 12-month dividend-growth nor dividend-price-ratio shocks materially influence returns, as their IRFs remain close to zero, confirming their statistical insignificance in return forecasting. Dividend growth itself reacts only weakly to return shocks, with small and short-lived movements that quickly vanish. Its only meaningful response arises from its own innovations, which produce a sharp but very short-lived spike that dissipates within a few months, illustrating the transitory and noisy nature of dividend-growth dynamics. Finally, valuation shocks lead to a brief negative impact on trailing 12-month dividend growth, which rapidly returns to its baseline, further reinforcing the weak cross-variable transmission mechanisms in the system.

Now moving the core of our analysis, compared to returns and dividends, dividend-price ratio exhibits the most persistent and economically interpretable responses in the system. A return shock produces a short-run positive effect on dp_t, which gradually declines to zero over the medium horizon. This response mirrors the estimated dynamics in which higher recent returns initially compress the valuation ratio (via rising prices) but subsequently give rise to corrective adjustments. A shock to cumulate dividend growth triggers a large and immediate increase in the dividend-price ratio, rising close to 0.20 on impact before declining steadily over the next 20 months. This pattern reflects the mechanical identity $dp_t = d_t - p_t$, where an unexpected surge in trailing 12-month dividend growth boosts the dividend component relative to prices. Finally, a shock to the dividend–price ratio generates a very persistent positive response, remaining well above zero throughout the entire horizon. The slow decay mirrors the highly persistent coefficient on dp_tin the VAR estimates, above 1.03 at lag 1, indicating that valuation ratios revert only gradually.

<u>In-sample vs Out-sample performance</u>

In terms of performance, our PVM-based model exhibits markedly different behaviour across the in-sample (2002/01/01 – 2021/12/01) and out-of-sample periods (2022/01/01 – 2025/11/01). In-sample, the model achieves a correlation of 0.84 between actual and model-implied prices, but its RMSE of 41.49 and high MAPE of 76.7% indicate substantial level deviations, reflected also in a very low pseudo-R^2 of 0.015. This pattern suggests that, while the VAR-based reconstruction tracks broad co-movements in prices during estimation, it struggles to match their scale, likely due to noise in dividend-growth dynamics and the strong persistence of valuation ratios. Out-of-sample performance improves sharply: the RMSE falls to 22.00 and the MAPE drops to 9.46%, while the correlation rises to 0.95. Most notably, the pseudo-R^2 increases to 0.893, indicating that the PVM-implied prices explain almost 90% of the variation in observed prices during the evaluation period.

<u>Out-of-sample performance - comparison to the martingale benchmark</u>

Now comparing the out-of-sample performance of the present-value model (PVM) with the martingale benchmark, our analysis reveals a clear horizon-dependent pattern. At short horizons, particularly at lag 1, the martingale performs substantially better, with an RMSE of 13.00, a MAPE of 5.36%, and a pseudo-R^2 of 0.963, far exceeding the PVM's RMSE of 21.99, MAPE of 9.46%, and pseudo-R^2 of 0.893 (Figure 18). However, with tried to investigate at which lag our PVM-based model overperforms martingale evolutions in predicting prices. We observe that his ranking reverses at longer horizons. By lag 4, the martingale's accuracy deteriorates sharply (RMSE rises to 29.50, MAPE to 11.12%, and pseudo-R^2 falls to 0.812), whereas the PVM remains stable and begins to outperform it (Figure 22). This switch occurs because the martingale is optimal over very short horizons, where price changes behave nearly unpredictably and resemble a random walk, but rapidly accumulates error when extended forward, as it assumes no mean reversion or valuation adjustment. In contrast, the PVM embeds slow-moving dividend and valuation dynamics through the VAR(2) structure, allowing it to capture medium-run predictability that becomes increasingly relevant at multi-period horizons. As a result, while the martingale dominates at the 1-step horizon, the PVM overtakes it by lag 4 (Figure 22), reflecting the model's strength in capturing the longer-run present-value structure of synthetic equity prices.

This horizon-dependent pattern aligns closely with the broader empirical literature on return and equity-premium predictability. At short horizons, our results show that the martingale benchmark dominates decisively: at lag 1 it delivers an RMSE of 13.00, a MAPE of 5.36%, and a pseudo-R^2 of 0.963, compared with the PVM's RMSE of 21.99, MAPE of 9.46%, and pseudo-R^2 of 0.893 (Figure 18). This finding is consistent with well-known real-time forecasting results, such as in Goyal and Welch (2008), showing that simple historical-mean or random-walk benchmarks often outperform more sophisticated predictive systems, even when valuation ratios exhibit statistically significant in-sample predictive power. From the perspective of the Efficient Markets Hypothesis (Fama, 1970), this is not surprising: price changes at high frequency tend to resemble a near-martingale, making short-horizon forecasts extremely difficult for any structural model.

Where the PVM gains its comparative advantage is at longer horizons, and this becomes visible around lag 4. The martingale, optimal in the short run precisely because it assumes no predictable structure, accumulates error rapidly as the forecast horizon extends. Its lack of mean reversion, its inability to incorporate slow-moving valuation adjustments, and its exclusion of dividend-growth information cause its accuracy to deteriorate sharply by lag 4. In contrast, the PVM embeds the persistent dynamics of dividends and the dividend, price ratio through the VAR(2) structure. These slow-moving components become increasingly relevant at medium-run horizons, allowing the PVM to exploit predictable variation in discount rates and cash-flow expectations that the martingale ignores. As these fundamental adjustments compound through time, the PVM's relative accuracy improves and eventually surpasses the martingale by lag 4, reflecting the model's ability to capture the long-run present-value structure inherent in the synthetic price data.

5. Conclusion

In this project we confronted two competing descriptions of assets price behavior: the martingale view implied by the efficient market hypothesis and a variant of the fundamental present value model under theoretical framework of Campbell, Shiller, and Cochrane. Using the synthesized Index and real-world Nasdaq 100, hourly, daily, weekly and monthly data we found that prices at very short horizons behave close to a martingale difference process, drift being the negligible part, sign changes are approximately independent, and log prices are well described by unit-root dynamics. However, as the horizon lengthens, clear and systematic deviations emerge returns display a positive drift, some serial dependence and mean reversion, so that prices resemble a submartingale with a deterministic growth component rather than a strict martingale. The martingale framework therefore remains a useful benchmark, especially at high frequencies, but it does not fully capture medium- and long-horizon dynamics. When looking at our VAR model PVM analysis the previous conclusion is reinforced. In-sample, return predictability arises almost entirely from autoregressive structure, whereas dividend growth is extremely noisy and valuation ratios are highly persistent. Out-of-sample, the PVM performs poorly at very short horizons, where the martingale dominates, but begins to outperform the martingale at longer horizons (around four periods ahead), precisely because it embeds slow-moving dividend and valuation dynamics that accumulate predictability over time. This horizon-dependent reversal mirrors the broader empirical literature: short-run equity returns are nearly unpredictable, but medium-run variation in discount rates and valuation ratios generates mild but systematic predictability. Taken together, this shows that both models capture important but distinct aspects of price dynamics: the martingale accurately describes high-frequency behaviour, while the fundamental present-value structure becomes increasingly relevant as the horizon expands. At short horizons, stock prices behave very much like a martingale; at longer horizons, a fundamental present-value model provides a more accurate description of their dynamics and predictive structure.

6. Bibliography

Malkiel, B. G. (1989). Efficient market hypothesis. In J. Eatwell, M. Milgate, & P. Newman (Eds.), Finance (pp. xx–xx). Palgrave Macmillan. https://doi.org/10.1007/978-1-349-20213-3_13

Roberts, H. (1967). Statistical versus clinical prediction of the stock market. Center for Research in Security Prices, University of Chicago.

Samuelson, P. A. (1973). Proof that properly discounted present values of assets vibrate randomly. The Bell Journal of Economics and Management Science, 4(2), 369–374. https://doi.org/10.2307/3003046

Cowles, A., III, & Jones, H. E. (1937). Some a posteriori probabilities in stock market action. Econometrica, 5(3), 280–294. https://doi.org/10.2307/1907031

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. The Journal of Finance, 25(2), 383–417. https://www.jstor.org/stable/2325486

Cochrane, J. H. (1992). Explaining the variance of price–dividend ratios. The Review of Financial Studies, 5(2), 243–280.

Cochrane, J. H. (2008). The Dog That Did Not Bark: A Defense of Return Predictability. Review of Financial Studies, 21(4), 1533–1575
1 Golez, B., & Koudijs, P. (2015). Four Centuries of Return Predictability. NBER Working Paper No. 20814.

Goyal, A., & Welch, I. (2008). A comprehensive look at the empirical performance of equity premium prediction. The Review of Financial Studies, 21(4), 1455–1508. https://doi.org/10.1093/rfs/hhm014

7. Appendix

Table 1: Descriptive Statistics of Log Returns and Log Prices

	Log Returns		Log Prices	
	NASDAQ	Synthetic Nasdaq	NASDAQ	Synthetic Nasdaq
Hourly				
count	1528	–	1528	–
mean	0.000112	–	10.002156	–
std	0.005451	–	0.094132	–
min	−0.042953	–	9.745003	–
25%	−0.001446	–	9.948075	–
50%	0.000181	–	9.994071	–
75%	0.001763	–	10.073058	–
max	0.052267	–	10.171774	–
Daily				
count	6007	8719	6007	8719
mean	0.000457	0.000458	8.276552	3.468807
std	0.014742	0.012067	0.928151	1.279655
min	−0.130032	−0.134961	6.690395	1.442647
25%	−0.005949	−0.002139	7.457430	2.317325
50%	0.001095	0.000000	8.151545	3.450479
75%	0.007674	0.004450	8.993174	4.503796
max	0.118493	0.127330	10.170451	5.956377
Weekly				
count	1245	1245	1246	1246
mean	0.002184	0.003218	8.277815	3.467887
std	0.031041	0.029130	0.929460	1.279970
min	−0.145331	−0.142071	6.690395	1.466157
25%	−0.013546	−0.011079	7.456888	2.313251
50%	0.004457	0.005062	8.153410	3.450479
75%	0.019467	0.019346	9.000040	4.503838
max	0.131928	0.146135	10.164810	5.949405
Monthly				
count	286	286	286	286
mean	0.009671	0.014050	8.283990	3.466151
std	0.055780	0.055008	0.934406	1.282277
min	−0.177875	−0.175058	6.724457	1.512968
25%	−0.020609	−0.017959	7.459245	2.303929
50%	0.016548	0.019131	8.166762	3.454316
75%	0.047857	0.049794	9.017023	4.496951
max	0.172783	0.155762	10.160380	5.927630

Table 2: Descriptive statistics of dividend yields and risk-free rate (log-levels and log-returns)

Series	Freq.	Count	Mean	Std.	Min	25%	50%	75%	Max
Dividend log-levels									
div_daily_log	Daily	2382	-6.519079	1.778922	-11.281414	-7.721611	-6.670531	-5.279619	-0.801863
div_weekly_log	Weekly	1066	-5.564421	1.886488	-11.281414	-6.665165	-5.528418	-4.179037	-0.800633
div_monthly_log	Monthly	287	-3.864623	1.628998	-7.871917	-4.860427	-3.750157	-2.689579	-0.790764
Dividend log-returns									
div_daily_log_ret	Daily	955	0.011352	2.302240	-6.632168	-1.504047	0.167473	1.610484	7.381785
div_weekly_log_ret	Weekly	921	0.039192	1.943806	-6.984983	-1.188176	0.199333	1.257762	7.157157
div_monthly_log_ret	Monthly	286	0.017768	1.838711	-4.206923	-1.307986	-0.803297	1.733723	6.924939
Risk-free log-levels									
rf_daily_log	Daily	8582	-0.744516	2.011548	-4.605170	-2.525729	-0.030459	0.973614	1.795087
rf_weekly_log	Weekly	1229	-0.709403	1.975599	-4.605170	-2.525729	-0.030459	0.993252	1.735189
rf_monthly_log	Monthly	284	-0.769362	2.036126	-4.605170	-2.525729	-0.035641	0.982058	1.722767
Risk-free log-returns									
rf_daily_log_ret	Daily	8541	-0.000089	0.192716	-2.639057	-0.001998	0.000000	0.000000	1.945910
rf_weekly_log_ret	Weekly	1221	0.000709	0.353999	-2.484907	-0.028492	0.000000	0.037622	2.397895
rf_monthly_log_ret	Monthly	281	0.003119	0.594933	-3.878121	-0.090384	0.000000	0.130712	2.944439

Table 3: Martingale / Submartingale Drift Tests for NASDAQ and Synthetic Log Returns

	Log Returns	
	NASDAQ	Synthetic Nasdaq
Hourly		
Number of obs	1528	–
Mean return	1.116660	–
Std dev	5.450903	–
Std error of mean	1.394462	–
t-statistic	0.801000	–
Two-sided p-value (H1: $\mu \neq 0$)	0.423400	–
One-sided p-value (H1: $\mu < 0$)	0.788300	–
One-sided p-value (H1: $\mu > 0$)	0.211700	–
Interpretation	MARTINGALE	–
Daily		
Number of obs	6007	8719
Mean return	4.565888	4.578064
Std dev	1.474215	1.206679
Std error of mean	1.902094	1.292286
t-statistic	2.400000	3.543000
Two-sided p-value (H1: $\mu \neq 0$)	0.016400	0.000398
One-sided p-value (H1: $\mu < 0$)	0.991800	0.999800
One-sided p-value (H1: $\mu > 0$)	0.008202	0.000199
Interpretation	SUBMARTINGALE	SUBMARTINGALE
Weekly		
Number of obs	1245	1245
Mean return	2.183561	3.217700
Std dev	3.104117	2.912954
Std error of mean	8.797380	8.255606
t-statistic	2.482000	3.898000
Two-sided p-value (H1: $\mu \neq 0$)	0.013190	0.000102
One-sided p-value (H1: $\mu < 0$)	0.993400	0.999900
One-sided p-value (H1: $\mu > 0$)	0.006597	5.117000
Interpretation	SUBMARTINGALE	SUBMARTINGALE
Monthly		
Number of obs	286	286
Mean return	9.671425	1.405040
Std dev	5.577964	5.500801
Std error of mean	3.298319	3.252692
t-statistic	2.932000	4.320000
Two-sided p-value (H1: $\mu \neq 0$)	0.003638	2.161000
One-sided p-value (H1: $\mu < 0$)	0.998200	1
One-sided p-value (H1: $\mu > 0$)	0.001819	1.080000
Interpretation	SUBMARTINGALE	SUBMARTINGALE

Table 4: Cowles & Jones Runs Test for NASDAQ and Synthetic Log Prices

	Log Prices	
	NASDAQ	Synthetic Nasdaq
Hourly		
N_S (sequences, same sign)	756	–
N_R (reversals, opp. sign)	771	–
Total pairs N	1527	–
$\hat{C}_J = N_S/N_R$	0.980500	–
$\hat{p} = N_S/(N_S + N_R)$	0.495100	–
z-stat (H_0: $p = 0.5$)	-0.383900	–
p-value (two-sided)	0.701083	–
Interpretation	Cannot reject RW1 at 5%	–
Daily		
N_S (sequences, same sign)	2946	2289
N_R (reversals, opp. sign)	3056	2416
Total pairs N	6002	4705
$\hat{C}_J = N_S/N_R$	0.964000	0.947400
$\hat{p} = N_S/(N_S + N_R)$	0.490800	0.486500
z-stat (H_0: $p = 0.5$)	-1.419900	-1.851500
p-value (two-sided)	0.155649	0.064098
Interpretation	Cannot reject RW1 at 5%	Cannot reject RW1 at 5%
Weekly		
N_S (sequences, same sign)	642	646
N_R (reversals, opp. sign)	602	598
Total pairs N	1244	1244
$\hat{C}_J = N_S/N_R$	1.066400	1.080300
$\hat{p} = N_S/(N_S + N_R)$	0.516100	0.519300
z-stat (H_0: $p = 0.5$)	1.243900	2.310200
p-value (two-sided)	0.213524	0.020879
Interpretation	Cannot reject RW1 at 5%	Reject RW1 at 5%
Monthly		
N_S (sequences, same sign)	153	162
N_R (reversals, opp. sign)	132	123
Total pairs N	285	285
$\hat{C}_J = N_S/N_R$	1.159100	1.317100
$\hat{p} = N_S/(N_S + N_R)$	0.536800	0.568400
z-stat (H_0: $p = 0.5$)	1.243900	2.310200
p-value (two-sided)	0.213524	0.020879
Interpretation	Cannot reject RW1 at 5%	Reject RW1 at 5%

Table 5: One-Period Drift t-Tests for NASDAQ and Synthetic Log Returns

	Log Returns	
	NASDAQ	Synthetic Nasdaq
Hourly		
n	1528	–
Mean return	1.116660	–
Std dev	5.450903	–
t-statistic	0.800800	–
p-value (greater)	0.211691	–
Interpretation	No drift (fail to reject H_0)	–
Daily		
n	6007	8719
Mean return	4.565888	4.578064
Std dev	1.474215	1.206679
t-statistic	2.400500	3.542600
p-value (greater)	0.008202	0.000199
Interpretation	Reject H_0: drift	Reject H_0: drift
Weekly		
n	1245	1245
Mean return	2.183561	3.217700
Std dev	3.104117	2.912954
t-statistic	2.482100	3.897600
p-value (greater)	0.006597	0.000051
Interpretation	Reject H_0: drift	Reject H_0: drift
Monthly		
n	286	286
Mean return	9.671425	1.405040
Std dev	5.577964	5.500801
t-statistic	2.932200	4.319600
p-value (greater)	0.001819	0.000011
Interpretation	Reject H_0: drift	Reject H_0: drift

Table 6: Campbell Long-Horizon Drift Tests for NASDAQ and Synthetic Log Prices

	Log Prices	
	NASDAQ	Synthetic Nasdaq
Hourly, $h = 1$		
n_{blocks}	1529	–
Mean R^1	0.000000	–
Std R^1	0.000000	–
t-statistic	0.000000	–
p-value (greater)	0.500000	–
Interpretation	Cannot reject H_0	–
Hourly, $h = 5$		
n_{blocks}	305	–
Mean R^5	6.293712	–
Std R^5	1.050116	–
t-statistic	1.046700	–
p-value (greater)	0.148036	–
Interpretation	Cannot reject H_0	–
Hourly, $h = 20$		
n_{blocks}	76	–
Mean R^{20}	2.861457	–
Std R^{20}	2.125920	–
t-statistic	1.173400	–
p-value (greater)	0.122174	–
Interpretation	Cannot reject H_0	–
Hourly, $h = 60$		
n_{blocks}	25	–
Mean R^{60}	7.489696	–
Std R^{60}	3.019501	–
t-statistic	1.240200	–
p-value (greater)	0.113441	–
Interpretation	Cannot reject H_0	–
Daily, $h = 1$		
n_{blocks}	6008	8720
Mean R^1	0.000000	0.000000
Std R^1	0.000000	0.000000
t-statistic	0.000000	0.000000
p-value (greater)	0.500000	0.500000
Interpretation	Cannot reject H_0	Cannot reject H_0
Daily, $h = 5$		
n_{blocks}	1201	1744
Mean R^5	1.107302	2.004942
Std R^5	2.795648	2.201561
t-statistic	1.372600	3.803200
p-value (greater)	0.085061	0.000074

	Log Prices (continued)	
	NASDAQ	Synthetic Nasdaq
Interpretation	Cannot reject H_0	Reject H_0 (drift)
Daily, $h = 20$		
n_{blocks}	300	436
Mean R^{20}	8.025827	9.302401
Std R^{20}	5.639597	4.473682
t-statistic	2.464900	4.341800
p-value (greater)	0.007133	0.000009
Interpretation	Reject H_0 (drift)	Reject H_0 (drift)
Daily, $h = 60$		
n_{blocks}	100	145
Mean R^{60}	2.870699	2.860324
Std R^{60}	9.793853	7.349463
t-statistic	2.931100	4.686400
p-value (greater)	0.002097	0.000003
Interpretation	Reject H_0 (drift)	Reject H_0 (drift)
Weekly, $h = 1$		
n_{blocks}	1246	1246
Mean R^{1}	0.000000	0.000000
Std R^{1}	0.000000	0.000000
t-statistic	0.000000	0.000000
p-value (greater)	0.500000	0.500000
Interpretation	Cannot reject H_0	Cannot reject H_0
Weekly, $h = 5$		
n_{blocks}	249	249
Mean R^{5}	1.053369	1.174461
Std R^{5}	5.838298	5.468601
t-statistic	2.847000	3.388900
p-value (greater)	0.002391	0.000408
Interpretation	Reject H_0 (drift)	Reject H_0 (drift)
Weekly, $h = 20$		
n_{blocks}	62	62
Mean R^{20}	4.473148	6.546159
Std R^{20}	1.328890	1.148690
t-statistic	2.650500	4.487200
p-value (greater)	0.005113	0.000016
Interpretation	Reject H_0 (drift)	Reject H_0 (drift)
Weekly, $h = 60$		
n_{blocks}	20	20
Mean R^{60}	1.271727	1.881644
Std R^{60}	2.900463	2.646056
t-statistic	1.960800	3.180200
p-value (greater)	0.032362	0.002464
Interpretation	Reject H_0 (drift)	Reject H_0 (drift)

	Log Prices (continued)	
	NASDAQ	Synthetic Nasdaq
Monthly, $h = 1$		
n_{blocks}	287	287
Mean R^1	0.000000	0.000000
Std R^1	0.000000	0.000000
t-statistic	0.000000	0.000000
p-value (greater)	0.500000	0.500000
Interpretation	Cannot reject H_0	Cannot reject H_0
Monthly, $h = 5$		
n_{blocks}	57	57
Mean R^5	3.851290	4.068833
Std R^5	1.100881	1.330237
t-statistic	2.641200	2.309300
p-value (greater)	0.005343	0.012320
Interpretation	Reject H_0 (drift)	Reject H_0 (drift)
Monthly, $h = 20$		
n_{blocks}	14	14
Mean R^{20}	1.847369	2.365879
Std R^{20}	1.701330	2.129714
t-statistic	4.062800	4.156600
p-value (greater)	0.000672	0.000564
Interpretation	Reject H_0 (drift)	Reject H_0 (drift)
Monthly, $h = 60$		
n_{blocks}	4	4
Mean R^{60}	5.507973	8.172764
Std R^{60}	4.708712	4.453730
t-statistic	2.339500	3.670100
p-value (greater)	0.050635	0.017499
Interpretation	Cannot reject H_0	Reject H_0 (drift)

Table 7: RW2 DF Regressions and Key Statistics – Hourly Log Prices

NASDAQ				Synthetic
DF regression without trend				
Variable	Coef.	t	p	
const	0.0119	0.802	0.423	No synthetic hourly price series.
X_{t-1}	-0.0012	-0.795	0.427	
R^2	0.000			
N	1528			
DF regression with trend				
Variable	Coef.	t	p	
const	0.0496	2.101	0.036	No synthetic hourly price series.
trend	1.046×10^{-6}	2.050	0.041	
X_{t-1}	-0.0050	-2.103	0.036	
R^2	0.003			
N	1528			
Key DF statistics				
	No trend	With trend		
$\hat{\phi}$	-0.001178	-0.005031		No DF statistics for synthetic hourly prices.
s.e.($\hat{\phi}$)	0.001482	0.002393		
t-stat($\hat{\phi}$)	-0.7946	-2.1027		
p-value	0.426956	0.035654		
Decision	RW2	trend-stationary		
ADF lags (AIC-selected)		24		
ADF AIC		-11414.2601		

Table 8: RW2 DF Regressions and Key Statistics – Daily Log Prices

NASDAQ

DF regression without trend

Variable	Coef.	t	p
const	-0.0006	-0.353	0.724
X_{t-1}	0.0001	0.624	0.533
R^2			0.000
N			6007

DF regression with trend

Variable	Coef.	t	p
const	0.0212	3.482	0.001
trend	1.803×10^{-6}	3.731	0.000
X_{t-1}	-0.0032	-3.492	0.000
R^2			0.002
N			6007

Key DF statistics

	No trend	With trend
$\hat{\phi}$	0.000128	-0.003155
s.e.($\hat{\phi}$)	0.000205	0.000903
t-stat($\hat{\phi}$)	0.6239	-3.4923
p-value	0.532714	0.000482
Decision	RW2	trend-stationary

Synthetic

DF regression without trend

Variable	Coef.	t	p
const	0.0002	0.506	0.613
X_{t-1}	7.755×10^{-5}	0.768	0.443
R^2			0.000
N			8719

DF regression with trend

Variable	Coef.	t	p
const	0.0033	3.700	0.000
trend	1.3×10^{-6}	3.842	0.000
X_{t-1}	-0.0024	-3.681	0.000
R^2			0.002
N			8719

Key DF statistics

	No trend	With trend
$\hat{\phi}$	0.000078	-0.002449
s.e.($\hat{\phi}$)	0.000101	0.000665
t-stat($\hat{\phi}$)	0.7677	-3.6814
p-value	0.442687	0.000233
Decision	RW2	trend-stationary

Table 9: RW2 DF Regressions and Key Statistics – Weekly Log Prices

NASDAQ

DF regression without trend

Variable	Coef.	t	p
const	-0.0035	-0.441	0.660
X_{t-1}	0.0007	0.722	0.471
R^2			0.000
N			1245

DF regression with trend

Variable	Coef.	t	p
const	0.0969	3.473	0.001
trend	4.017×10^{-5}	3.749	0.000
X_{t-1}	-0.0145	-3.486	0.001
R^2			0.012
N			1245

Key DF statistics

	No trend	With trend
$\hat{\phi}$	0.000684	-0.014467
s.e.($\hat{\phi}$)	0.000948	0.004150
t-stat($\hat{\phi}$)	0.7215	-3.4861
p-value	0.470724	0.000507
Decision	RW2	trend-stationary

Synthetic

DF regression without trend

Variable	Coef.	t	p
const	0.0011	0.453	0.651
X_{t-1}	0.0006	0.954	0.340
R^2			0.001
N			1245

DF regression with trend

Variable	Coef.	t	p
const	0.0200	3.533	0.000
trend	5.549×10^{-5}	3.681	0.000
X_{t-1}	-0.0148	-3.493	0.000
R^2			0.012
N			1245

Key DF statistics

	No trend	With trend
$\hat{\phi}$	0.000616	-0.014807
s.e.($\hat{\phi}$)	0.000646	0.004239
t-stat($\hat{\phi}$)	0.9543	-3.4934
p-value	0.340097	0.000494
Decision	RW2	trend-stationary

Table 10: RW2 DF Regressions and Key Statistics – Monthly Log Prices

NASDAQ

DF regression without trend

Variable	Coef.	t	p
const	-0.0169	-0.572	0.568
X_{t-1}	0.0032	0.905	0.366
R^2			0.003
N			286

DF regression with trend

Variable	Coef.	t	p
const	0.3295	3.148	0.002
trend	0.0006	3.445	0.001
X_{t-1}	-0.0490	-3.151	0.002
R^2			0.043
N			286

Key DF statistics

	No trend	With trend
$\hat{\phi}$	0.003216	-0.049010
s.e.($\hat{\phi}$)	0.003555	0.015555
t-stat($\hat{\phi}$)	0.9046	-3.1507
p-value	0.366456	0.001803
Decision	RW2	trend-stationary

Synthetic

Variable	Coef.	t	p
const	0.0037	0.391	0.696
X_{t-1}	0.0030	1.176	0.241
R^2			0.005
N			286

Variable	Coef.	t	p
const	0.0731	3.348	0.001
trend	0.0009	3.508	0.001
X_{t-1}	-0.0534	-3.281	0.001
R^2			0.046
N			286

	No trend	With trend
$\hat{\phi}$	0.003000	-0.053353
s.e.($\hat{\phi}$)	0.002552	0.016260
t-stat($\hat{\phi}$)	1.1758	-3.2813
p-value	0.240645	0.001163
Decision	RW2	trend-stationary

Table 11: Autocorrelation and Ljung–Box Test Results

Dataset	**Obs.**	**LB Q(10)**	**p-value**
NASDAQ Hourly	1528	29.1328	0.001186
NASDAQ Daily	6007	89.1657	0.000000
NASDAQ Weekly	1245	12.0560	0.281325
NASDAQ Monthly	286	7.9244	0.636223
Synthetic Daily	8719	119.5110	0.000000
Synthetic Weekly	1245	21.5738	0.017429
Synthetic Monthly	286	6.2228	0.796214

Table 12: Variance Ratio Test Results (Lo–MacKinlay)

Dataset	**VR(2)**	**p(2)**	**VR(5)**	**p(5)**	**VR(10)**	**p(10)**
NASDAQ Hourly	1.0385	0.1319	1.0357	0.2025	0.9090	0.0016
NASDAQ Daily	0.9027	0.0000	0.8195	0.0000	0.7704	0.0000
NASDAQ Weekly	0.9382	0.0291	0.8599	0.0000	0.8350	0.0000
NASDAQ Monthly	1.0220	0.7095	1.0852	0.1883	0.9721	0.6751
Synthetic Daily	0.8980	0.0000	0.8487	0.0000	0.7935	0.0000
Synthetic Weekly	0.9159	0.0030	0.8442	0.0000	0.8417	0.0000
Synthetic Monthly	1.0390	0.5090	1.1062	0.1012	1.0692	0.2981

Table 13: OLS Regression Results

	Coef.	Std. Err.	t	P>—t—
const	5.6224	0.141	39.966	0.000
d	0.5533	0.034	16.095	0.000
Model Statistics				
R-squared				0.486
Adj. R-squared				0.484
F-statistic				259.1 (p = 1.75e-41)
No. Observations				276
AIC / BIC				729.8 / 737.0
Durbin–Watson				1.289

Table 14: Engle–Granger Residual ADF Test

ADF Statistic	0.487
p-value	0.984
Critical Values	
1% critical value	-3.455
5% critical value	-2.873
10% critical value	-2.573

Table 15: Campbell Approach - ADF Test on $dp_t = \log(D) - \log(P)$

ADF Statistic	-1.793
p-value	0.384
Critical Values	
1% critical value	-3.455
5% critical value	-2.873
10% critical value	-2.573

Table 16: OLS Regression Results: $\log(D_{12})$ on $\log(P)$

	Coef.	Std. Err.	t	P>—t—
const	4.4461	0.040	112.186	0.000
d12	1.1168	0.030	36.971	0.000
Model Statistics				
R-squared				0.833
Adj. R-squared				0.832
F-statistic				1367. (p = 1.70e-108)
No. Observations				276
AIC / BIC				419.4 / 426.7
Durbin–Watson				0.163

Table 17: Engle–Granger Residual ADF Test

ADF Statistic	-2.402
p-value	0.141
Critical Values	
1% critical value	-3.455
5% critical value	-2.873
10% critical value	-2.573

Table 18: ADF Test on $dp_t = \log(D12) - \log(P)$

ADF Statistic	-1.575
p-value	0.496
Critical Values	
1% critical value	-3.455
5% critical value	-2.873
10% critical value	-2.573

Table 19: Johansen Cointegration Test Results for $(p_t,\ d12_t)$

	Statistic	5% Crit.	1% Crit.
Trace Test			
$r = 0$	15.268	15.494	19.935
$r \leq 1$	0.111	3.842	6.635
Max-Eigenvalue Test			
$r = 0$	15.158	14.264	18.520
$r \leq 1$	0.111	3.842	6.635

Table 20: Estimated Cointegrating Vector

Raw eigenvector β	$[1.7163,\ -2.4949]$
Normalized on $p = 1$	$[1,\ -1.4536]$

Table 21: VAR Order Selection Criteria (* denotes minimum)

Lag	AIC	BIC	FPE	HQIC
0	1.931	1.976	6.896	1.949
1	-6.849	-6.670	0.001060	-6.777
2	-7.333	**-7.019***	0.0006538	**-7.206***
3	-7.292	-6.844	0.0006808	-7.112
4	-7.304	-6.721	0.0006732	-7.069
5	**-7.393***	-6.676	**0.0006159***	-7.104
6	-7.363	-6.511	0.0006352	-7.019
7	-7.311	-6.324	0.0006693	-6.913
8	-7.324	-6.203	0.0006613	-6.872
9	-7.277	-6.021	0.0006937	-6.771
10	-7.267	-5.877	0.0007017	-6.706

Table 22: Regression Results for Equation rx_{t+1}

Variable	Coeff.	Std. Error	t-Stat	p-Value
const	1.096921	0.691645	1.586	0.113
L1.rx$_{t+1}$	0.641838	0.062881	10.207	0.000
L1.Δd_{t+1}	-0.003526	0.039052	-0.090	0.928
L1.dp$_t$	0.617325	0.729207	0.847	0.397
L2.rx$_{t+1}$	0.318716	0.062874	5.069	0.000
L2.Δd_{t+1}	-0.521746	0.724606	-0.720	0.471
L2.dp$_t$	-0.475902	0.723602	-0.658	0.511

Table 23: Regression Results for Equation Δd_{t+1}

Variable	Coeff.	Std. Error	t-Stat	p-Value
const	-5.806475	0.951723	-6.101	0.000
L1.rx$_{t+1}$	0.139118	0.086526	1.608	0.108
L1.Δd_{t+1}	-1.202042	0.053736	-22.369	0.000
L1.dp$_t$	-0.231213	1.003409	-0.230	0.818
L2.rx$_{t+1}$	-0.109780	0.086516	-1.269	0.204
L2.Δd_{t+1}	-1.146174	0.997077	-1.150	0.250
L2.dp$_t$	-0.571760	0.995697	-0.574	0.566

Table 24: Regression Results for Equation dp_t

Variable	Coeff.	Std. Error	t-Stat	p-Value
const	-0.142869	0.061618	-2.319	0.020
L1.rx$_{t+1}$	0.012487	0.005602	2.229	0.026
L1.Δd_{t+1}	0.991003	0.003479	284.845	0.000
L1.dp$_t$	1.034988	0.064964	15.932	0.000
L2.rx$_{t+1}$	-0.015017	0.005601	-2.681	0.007
L2.Δd_{t+1}	-0.048665	0.064554	-0.754	0.451
L2.dp$_t$	-0.053281	0.064465	-0.827	0.409

Table 25. Correlation Matrix of Residuals

	rx_{t+1}	Δd_{t+1}	dp_t
rx_{t+1}	1.000000	-0.013657	0.196843
Δd_{t+1}	-0.013657	1.000000	-0.030429
dp_t	0.196843	-0.030429	1.000000

Table 26. Descriptive Statistics (Transposed)

Variable	count	mean	min	25%	50%	75%	max	std
P_t	273	71.039983	4.604349	11.205518	34.824557	94.677412	366.551887	84.654065
p_t	273	3.539590	1.527001	2.416406	3.550323	4.550475	5.904140	1.247178
D_t	273	0.673128	0.045918	0.200336	0.607705	0.962000	2.101338	0.535254
d_t	273	-0.802734	-3.080899	-1.607757	-0.498066	-0.038741	0.742574	1.006784
Rf_t	273	1.635311	0.010000	0.070000	0.890000	2.760000	5.600000	1.883840
rf_t	273	-0.857380	-4.605170	-2.659260	-0.116534	1.015231	1.722767	2.076096
r_{t+1}	273	0.031654	-0.154963	-0.000655	0.036594	0.066631	0.167075	0.054758
rx_{t+1}	273	0.884477	-1.766504	-0.970902	0.150184	2.688497	4.741796	2.080526
Δd_{t+1}	273	0.013639	-1.631618	0.000487	0.007250	0.018498	2.230354	0.181044
dp_t	273	-4.342324	-5.185934	-4.692070	-4.358479	-4.169771	-2.588515	0.524946
rx_{t+1}^{fit}	273	1.500195	-1.521613	-0.304281	2.162809	2.761494	4.610701	1.798430
Δd_{t+1}^{fit}	273	0.017931	-0.182369	0.009793	0.023412	0.035345	0.096233	0.038743
dp_t^{fit}	273	-4.270918	-5.166539	-4.465497	-4.339948	-4.180981	-2.577871	0.467091
D_t^{PVM}	273	0.515814	0.022964	0.039868	0.051283	0.192399	4.490413	1.050194
P_t^{PVM}	273	45.593992	0.344718	2.514459	3.808771	19.296353	333.891751	86.087905
Mispricing	273	25.445991	-42.787082	7.026660	14.687884	41.224926	101.242234	29.466747
Ratio	273	5.390850	-0.339368	0.652429	4.182819	9.130602	22.404041	5.181813

Table 27. In-Sample and Out-of-Sample PVM Performance

	RMSE	MAPE	$\mathrm{Corr}(\mathrm{P}_t, P_t^{PVM})$	Pseudo R^2
In-Sample (Train)	41.4856	76.72%	0.8398	0.0146
Out-of-Sample (Test)	21.9991	9.46%	0.9540	0.8926

Table 28. Out-of-Sample Martingale Performance (Optimized)

	RMSE	MAPE	$\mathrm{Corr}(\mathrm{P}_t, P_t^{mart})$	Pseudo R^2
Martingale (Optimal)	29.5014	11.12%	0.9331	0.8115

Table 29. Out-of-Sample Martingale Performance (Lag = 6)

	RMSE	MAPE	$\mathrm{Corr}(\mathrm{P}_t, P_t^{mart,6})$	Pseudo R^2
Lag = 6	37.9882	14.43%	0.8995	0.6832

Table 30. Out-of-Sample Martingale Performance (Lag = 12)

	RMSE	MAPE	$\mathrm{Corr}(\mathrm{P}_t, P_t^{mart,12})$	Pseudo R^2
Lag = 12	65.0093	23.89%	0.7937	-0.1610

<u>Figure 1</u>

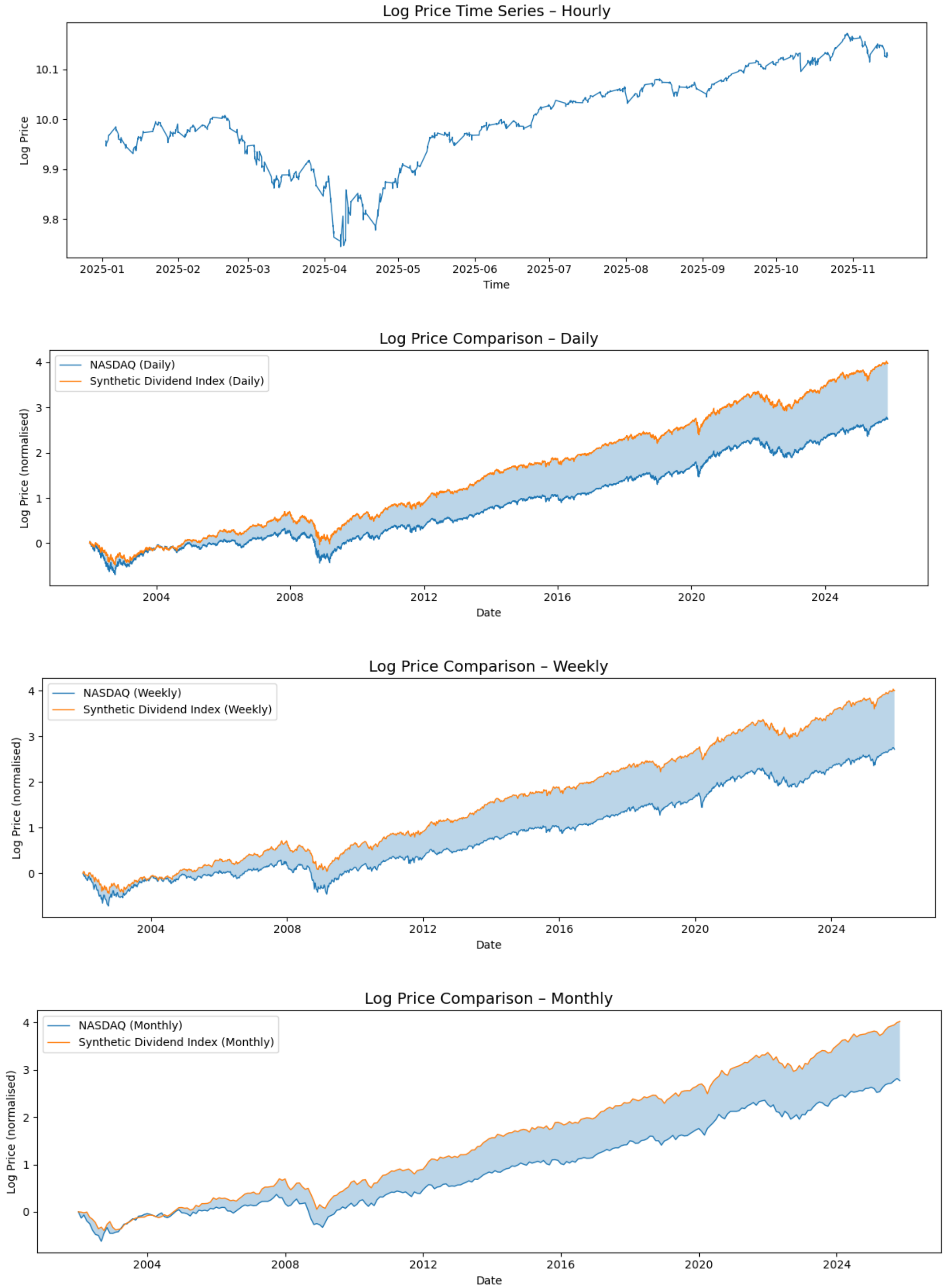
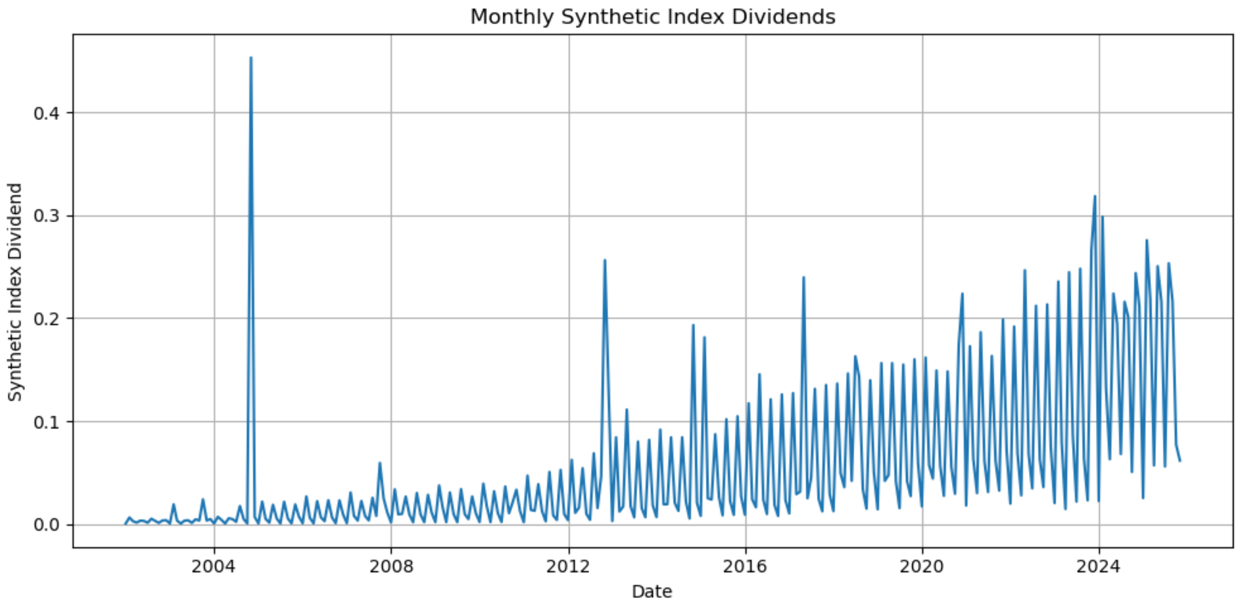

Log Price Time Series – Hourly
Log Price
10.1
10.0
9.9
9.8
2025-01
2025-02
2025-03
2025-04
2025-05
2025-06
2025-07
2025-08
2025-09
2025-10
2025-11
Time
Log Price Comparison – Daily
NASDAQ (Daily)
Synthetic Dividend Index (Daily)
Log Price (normalised)
4
3
2
1
0
2004
2008
2012
2016
2020
2024
Date
Log Price Comparison – Weekly
NASDAQ (Weekly)
Synthetic Dividend Index (Weekly)
Log Price (normalised)
4
3
2
1
0
2004
2008
2012
2016
2020
2024
Date
Log Price Comparison – Monthly
NASDAQ (Monthly)
Synthetic Dividend Index (Monthly)
Log Price (normalised)
4
3
2
1
0
2004
2008
2012
2016
2020
2024
Date

Figure 2


Monthly Synthetic Index Dividends
Synthetic Index Dividend
0.0
0.1
0.2
0.3
0.4
2004
2008
2012
2016
2020
2024
Date

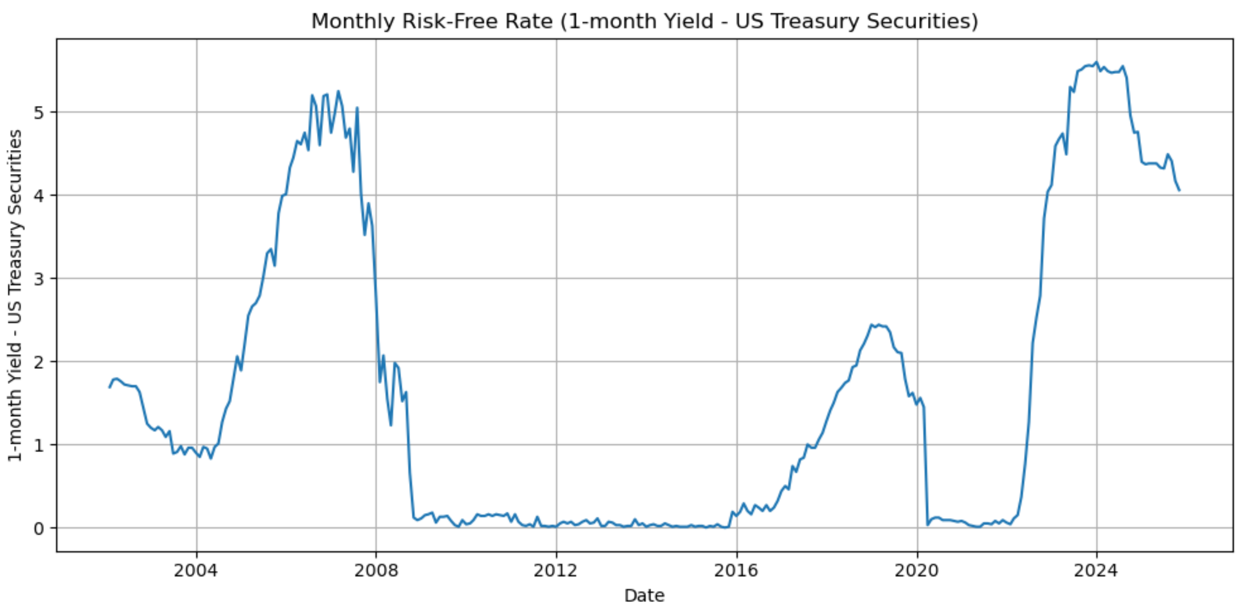
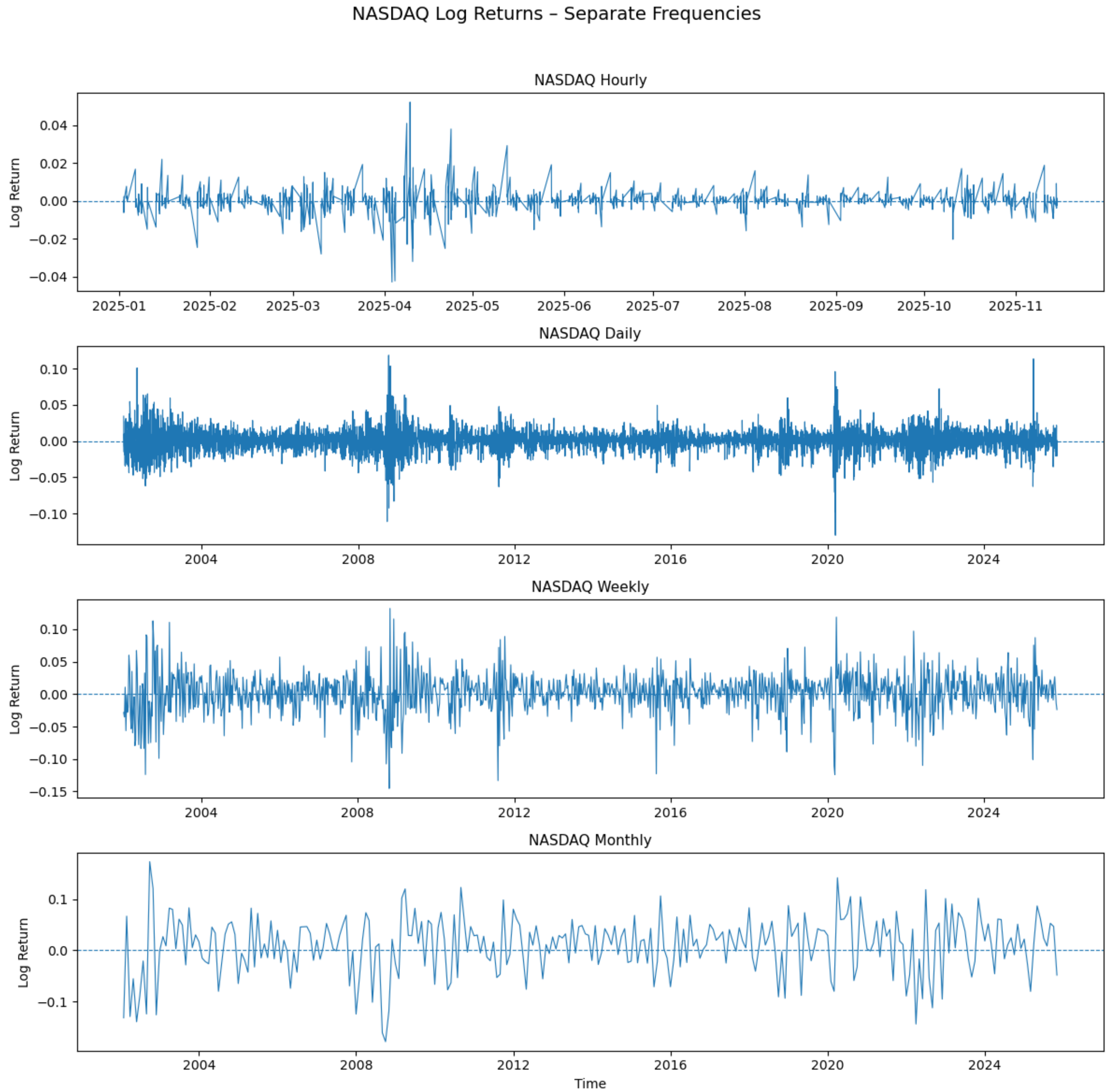

Monthly Risk-Free Rate (1-month Yield - US Treasury Securities)
1-month Yield - US Treasury Securities
0
1
2
3
4
5
2004
2008
2012
2016
2020
2024
Date

Figure 3


NASDAQ Log Returns – Separate Frequencies
NASDAQ Hourly
Log Return
0.04
0.02
0.00
−0.02
−0.04
2025-01
2025-02
2025-03
2025-04
2025-05
2025-06
2025-07
2025-08
2025-09
2025-10
2025-11
NASDAQ Daily
0.10
0.05
0.00
−0.05
−0.10
2004
2008
2012
2016
2020
2024
NASDAQ Weekly
0.10
0.05
0.00
−0.05
−0.10
−0.15
NASDAQ Monthly
0.1
0.0
−0.1
Time

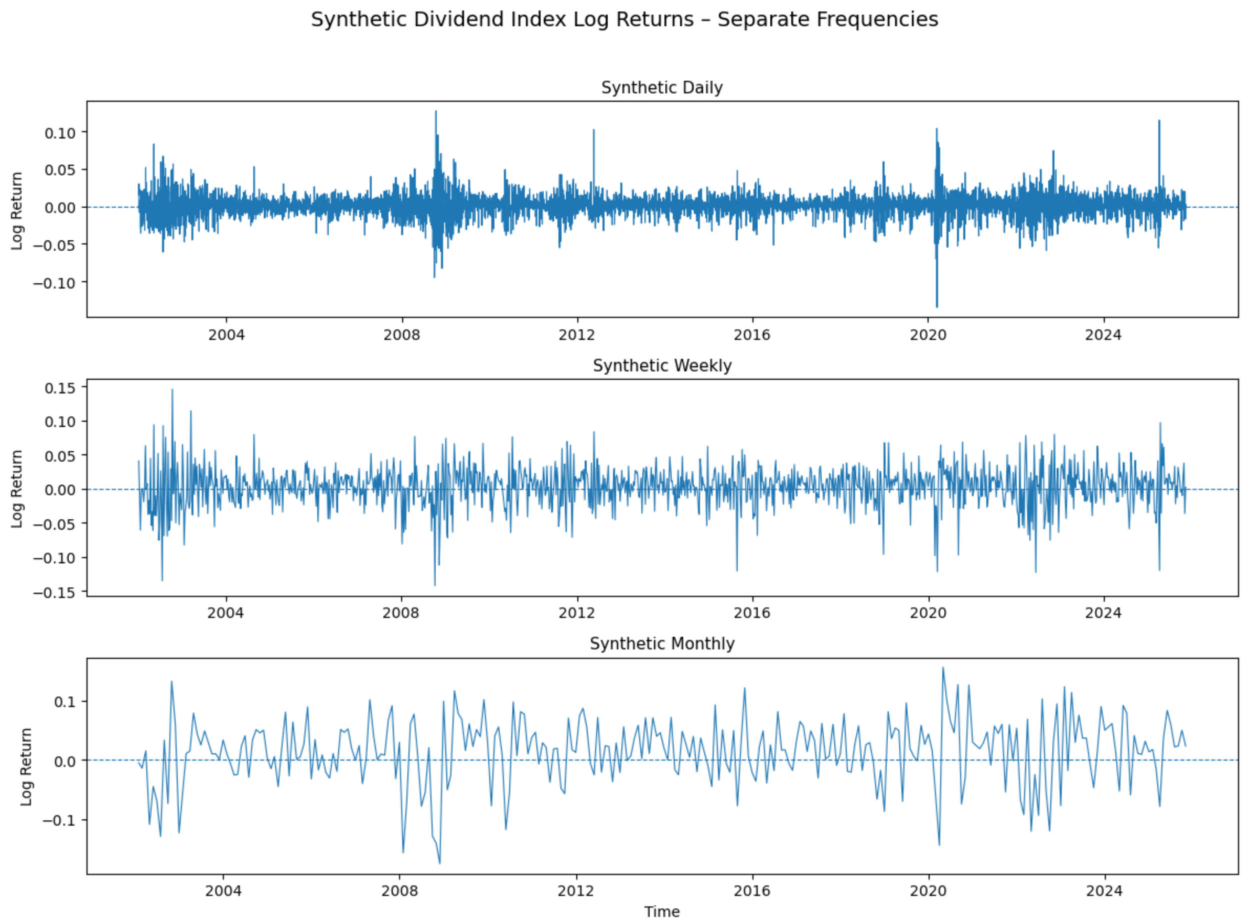
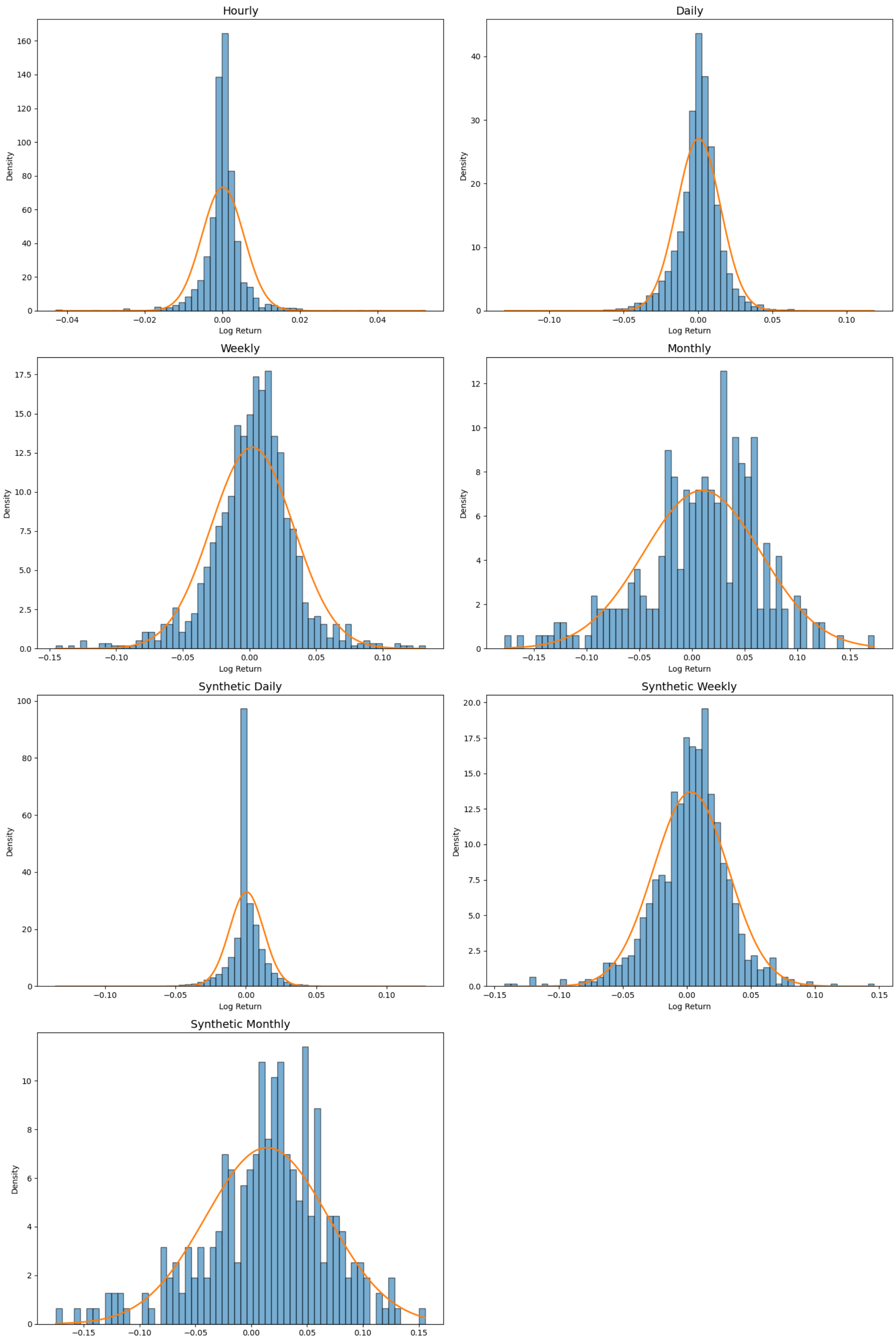
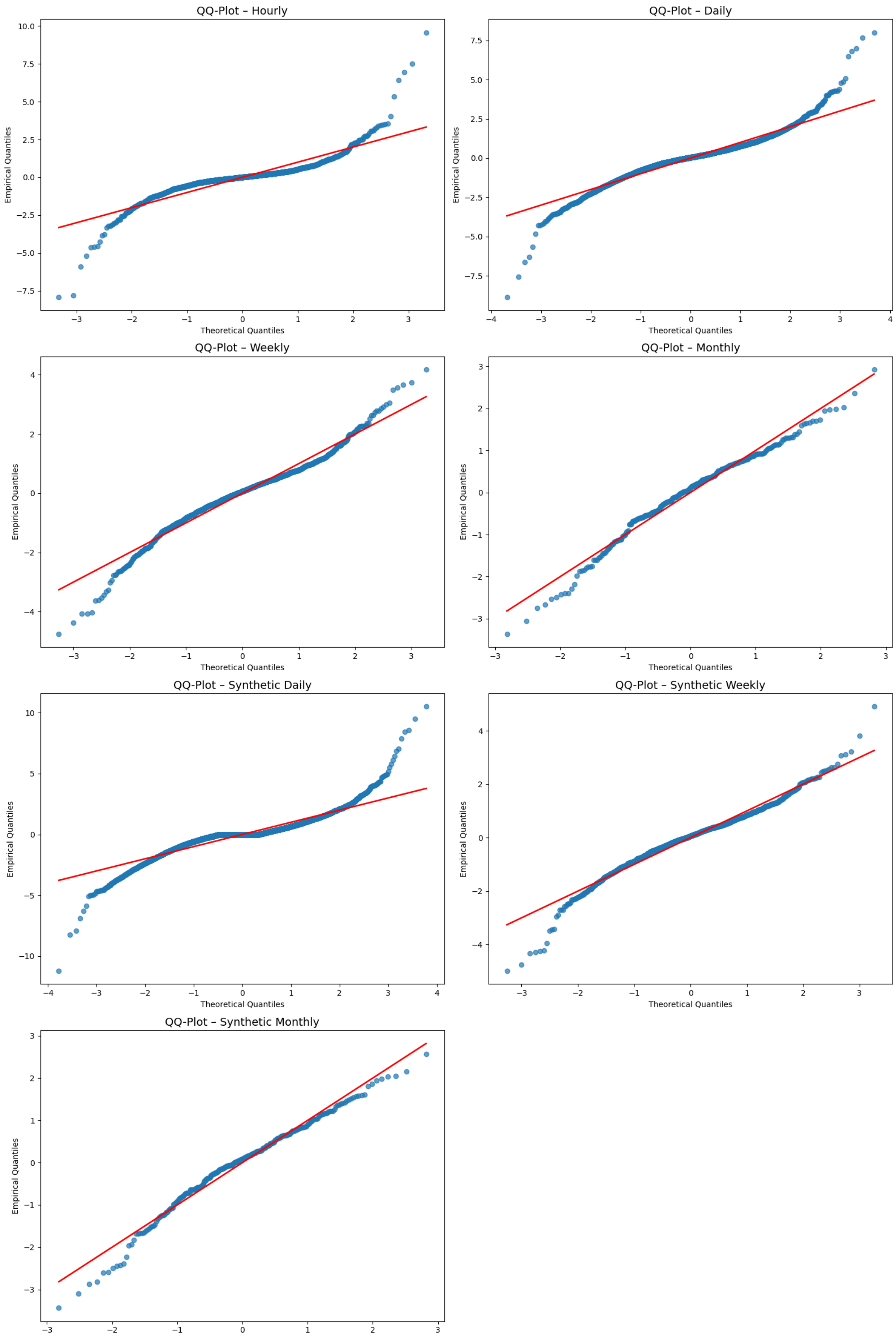

Synthetic Dividend Index Log Returns – Separate Frequencies
Synthetic Daily
Log Return
0.10
0.05
0.00
−0.05
−0.10
2004
2008
2012
2016
2020
2024
Synthetic Weekly
Log Return
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
2004
2008
2012
2016
2020
2024
Synthetic Monthly
Log Return
0.1
0.0
−0.1
2004
2008
2012
2016
2020
2024
Time

Figure 4



Hourly
Daily
Weekly
Monthly
Synthetic Daily
Synthetic Weekly
Synthetic Monthly
Density
Log Return

Figure 5



QQ-Plot – Hourly
QQ-Plot – Daily
QQ-Plot – Weekly
QQ-Plot – Monthly
QQ-Plot – Synthetic Daily
QQ-Plot – Synthetic Weekly
QQ-Plot – Synthetic Monthly
Empirical Quantiles
Theoretical Quantiles

Figure 6

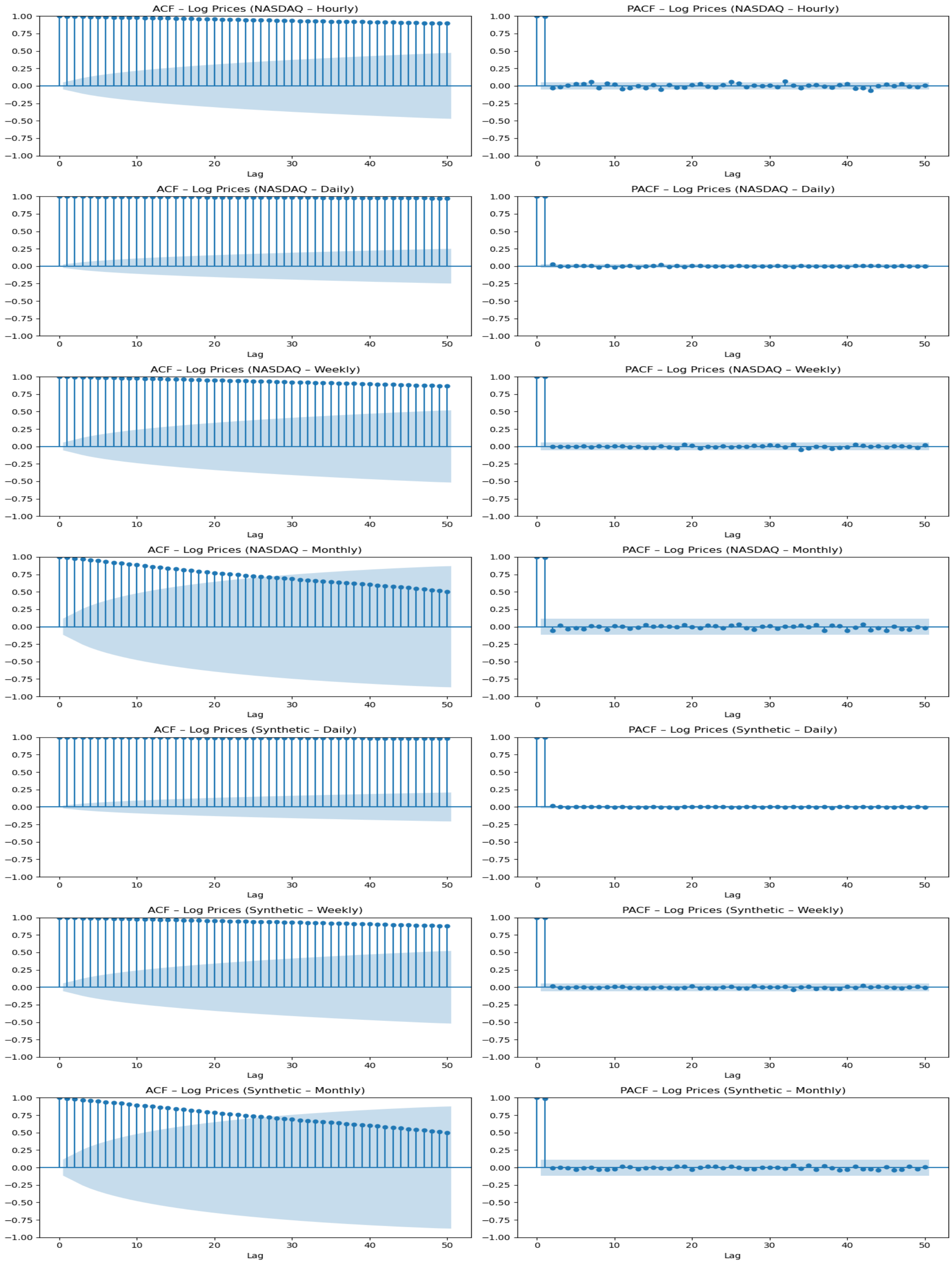

ACF – Log Prices (NASDAQ – Hourly)
PACF – Log Prices (NASDAQ – Hourly)
ACF – Log Prices (NASDAQ – Daily)
PACF – Log Prices (NASDAQ – Daily)
ACF – Log Prices (NASDAQ – Weekly)
PACF – Log Prices (NASDAQ – Weekly)
ACF – Log Prices (NASDAQ – Monthly)
PACF – Log Prices (NASDAQ – Monthly)
ACF – Log Prices (Synthetic – Daily)
PACF – Log Prices (Synthetic – Daily)
ACF – Log Prices (Synthetic – Weekly)
PACF – Log Prices (Synthetic – Weekly)
ACF – Log Prices (Synthetic – Monthly)
PACF – Log Prices (Synthetic – Monthly)
Lag

Figure 7

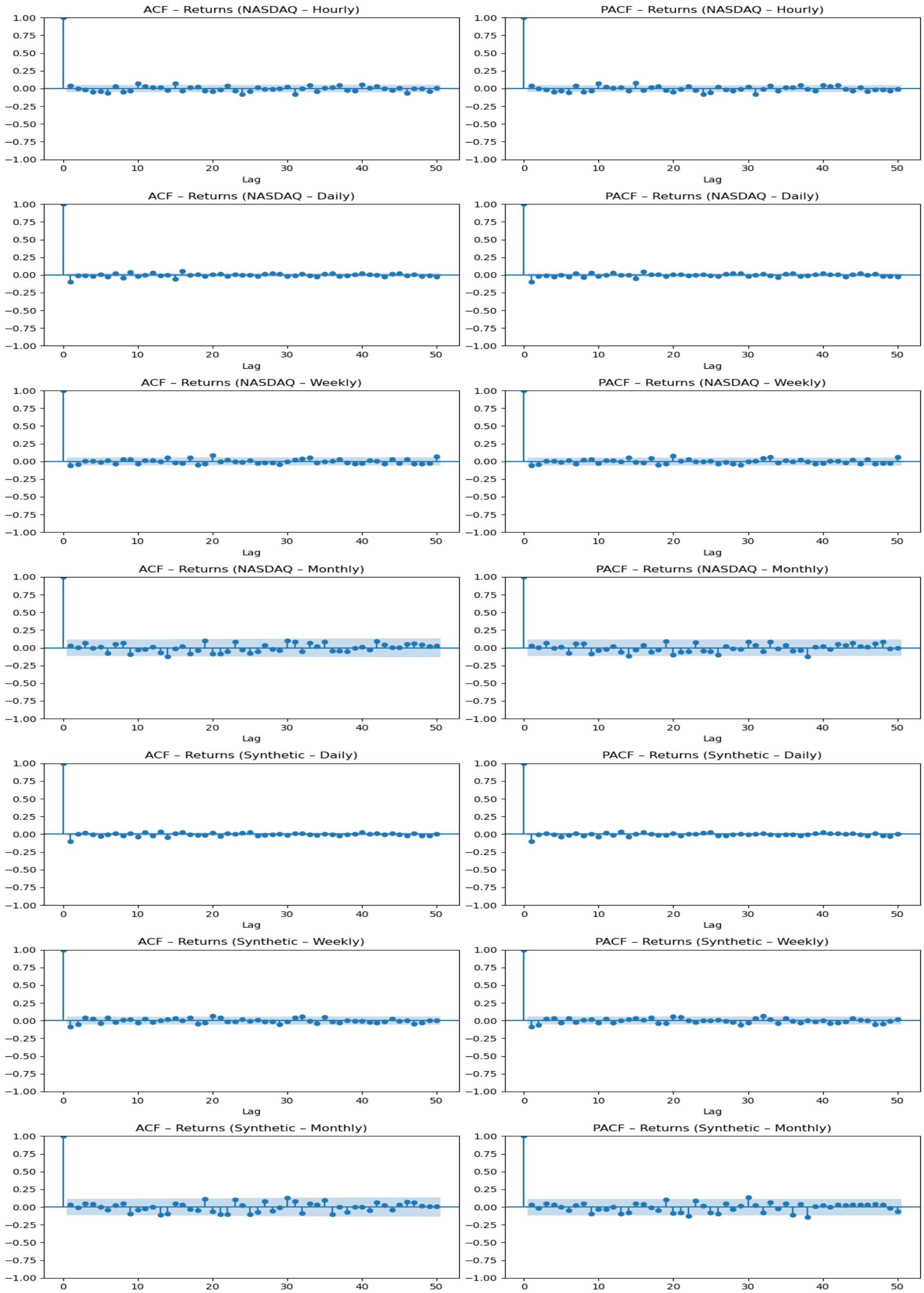
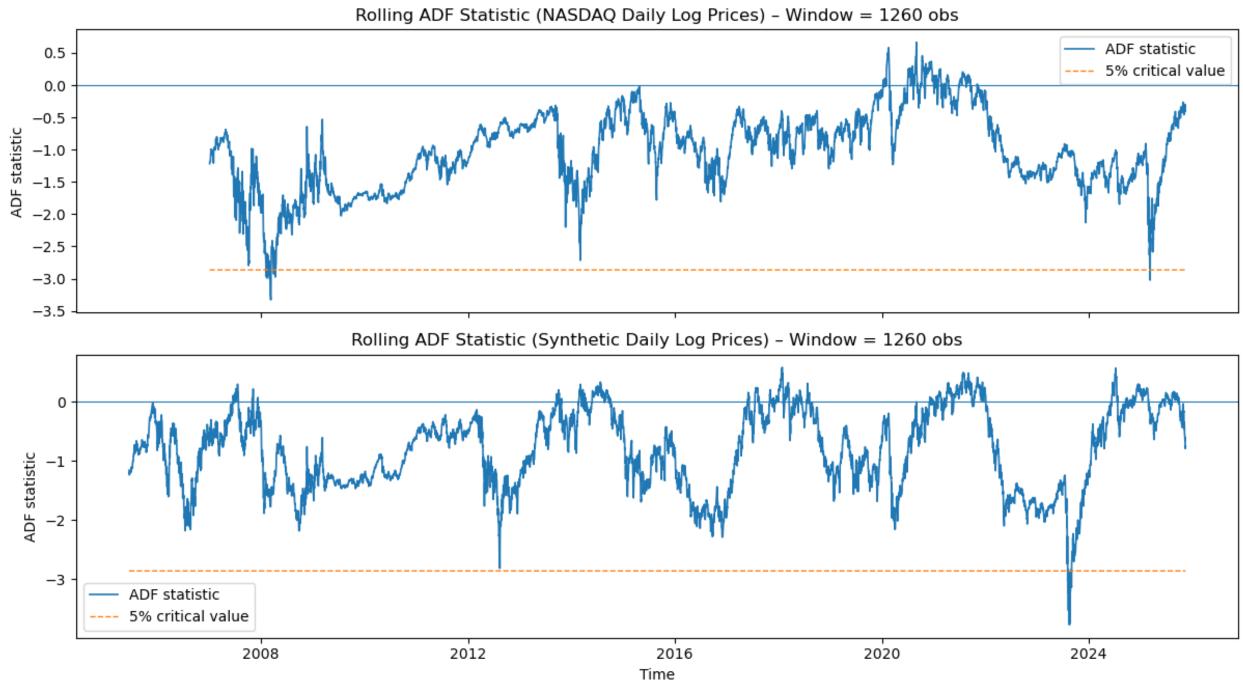

ACF – Returns (NASDAQ – Hourly)
PACF – Returns (NASDAQ – Hourly)
ACF – Returns (NASDAQ – Daily)
PACF – Returns (NASDAQ – Daily)
ACF – Returns (NASDAQ – Weekly)
PACF – Returns (NASDAQ – Weekly)
ACF – Returns (NASDAQ – Monthly)
PACF – Returns (NASDAQ – Monthly)
ACF – Returns (Synthetic – Daily)
PACF – Returns (Synthetic – Daily)
ACF – Returns (Synthetic – Weekly)
PACF – Returns (Synthetic – Weekly)
ACF – Returns (Synthetic – Monthly)
PACF – Returns (Synthetic – Monthly)
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
10
20
30
40
50
Lag

Figure 8



Rolling ADF Statistic (NASDAQ Daily Log Prices) – Window = 1260 obs
ADF statistic
5% critical value
Rolling ADF Statistic (Synthetic Daily Log Prices) – Window = 1260 obs
ADF statistic
5% critical value
2008
2012
2016
2020
2024
Time

Figure 9

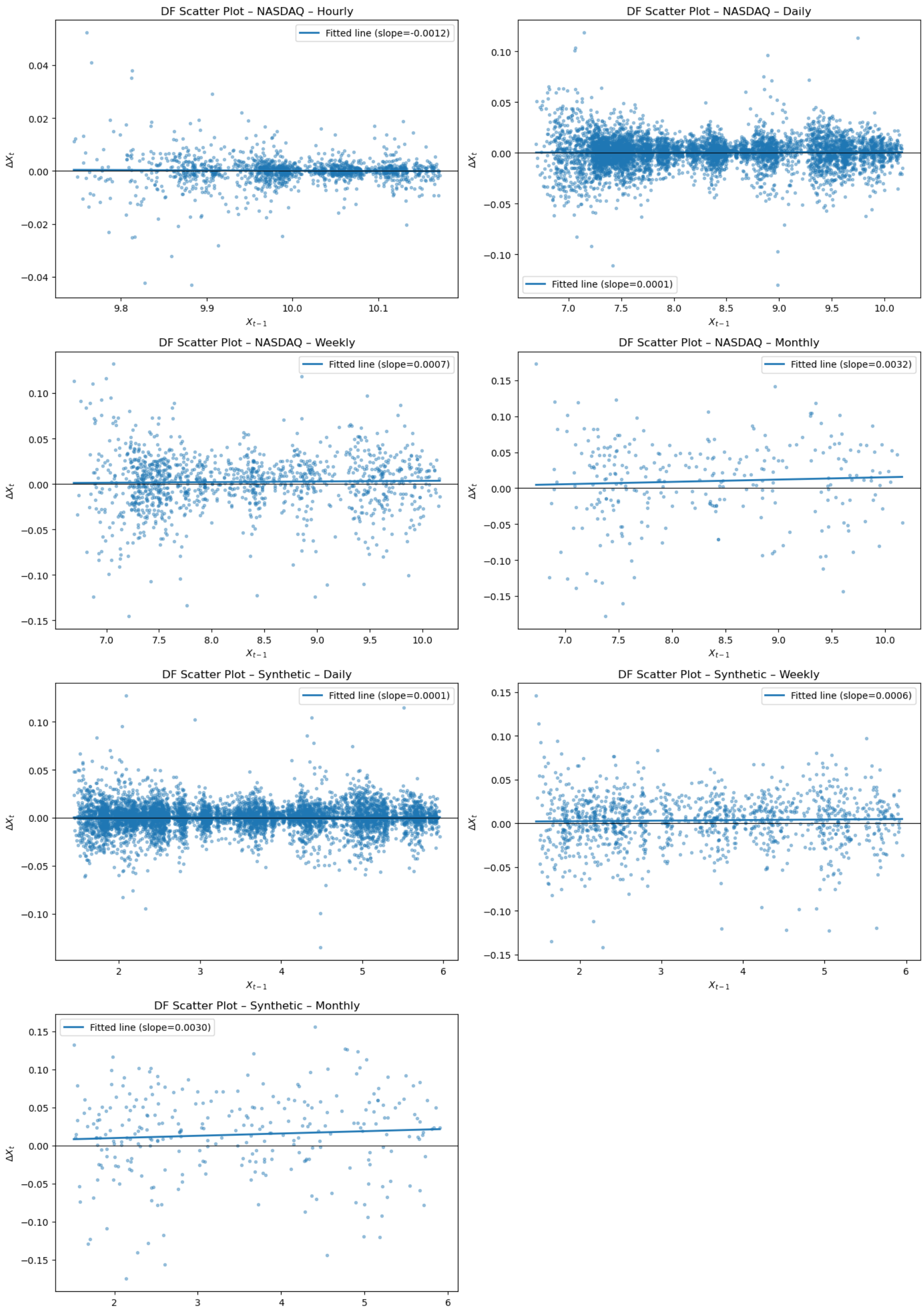
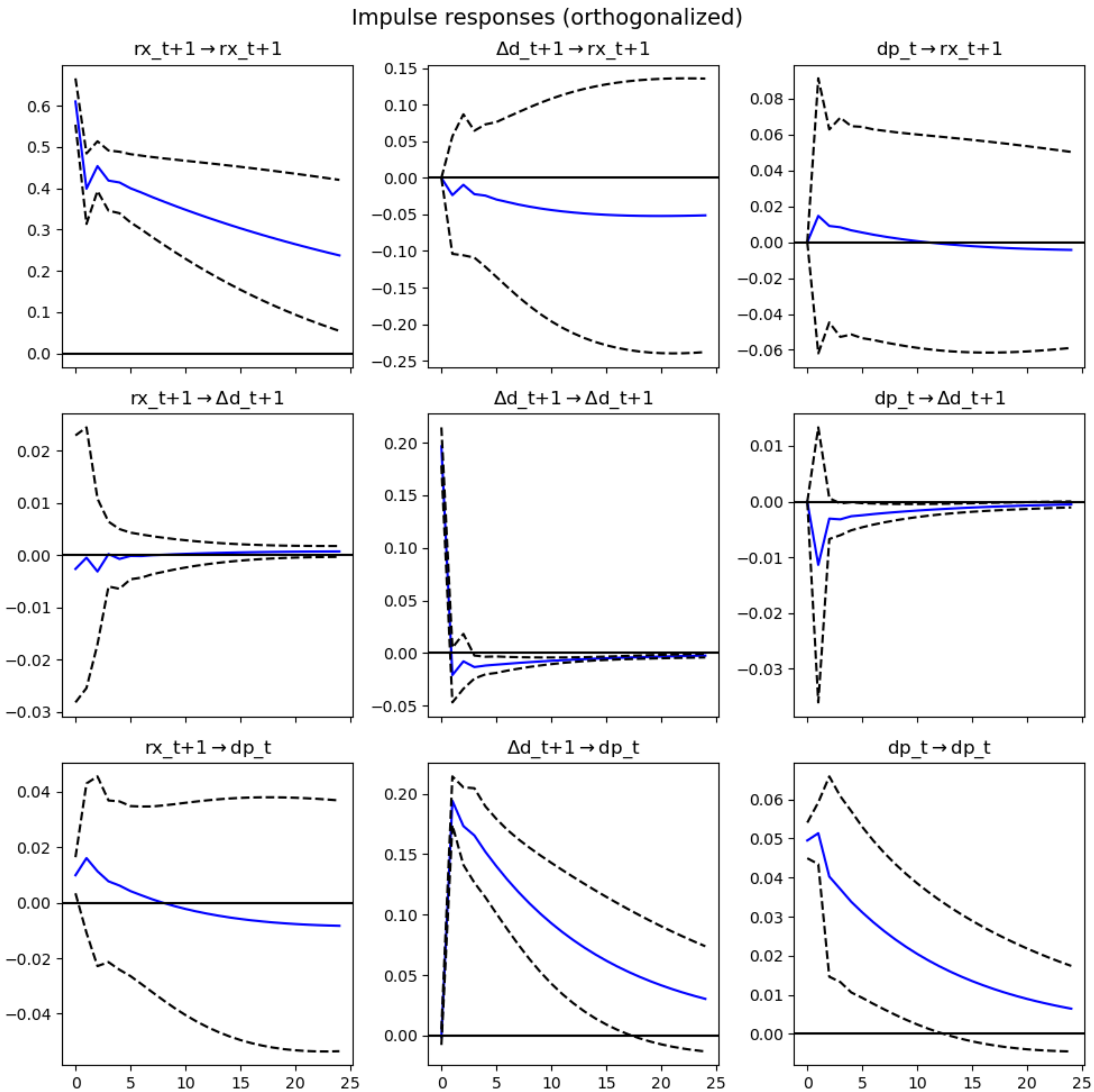

DF Scatter Plot – NASDAQ – Hourly
Fitted line (slope=-0.0012)
DF Scatter Plot – NASDAQ – Daily
Fitted line (slope=0.0001)
DF Scatter Plot – NASDAQ – Weekly
Fitted line (slope=0.0007)
DF Scatter Plot – NASDAQ – Monthly
Fitted line (slope=0.0032)
DF Scatter Plot – Synthetic – Daily
Fitted line (slope=0.0001)
DF Scatter Plot – Synthetic – Weekly
Fitted line (slope=0.0006)
DF Scatter Plot – Synthetic – Monthly
Fitted line (slope=0.0030)
ΔX_t
X_{t-1}

Figure 10



Impulse responses (orthogonalized)
rx_t+1 → rx_t+1
Δd_t+1 → rx_t+1
dp_t → rx_t+1
rx_t+1 → Δd_t+1
Δd_t+1 → Δd_t+1
dp_t → Δd_t+1
rx_t+1 → dp_t
Δd_t+1 → dp_t
dp_t → dp_t

Figure 11

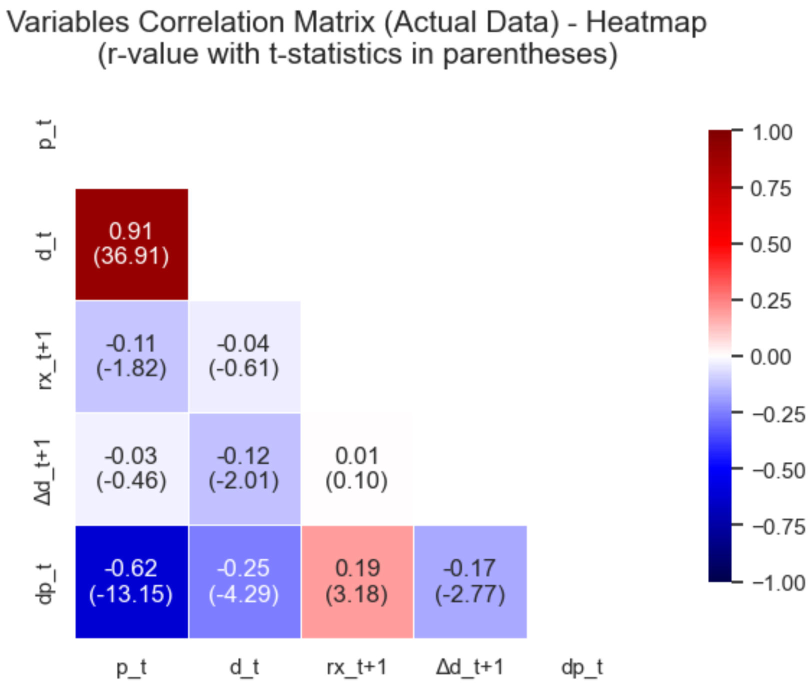


Figure 12

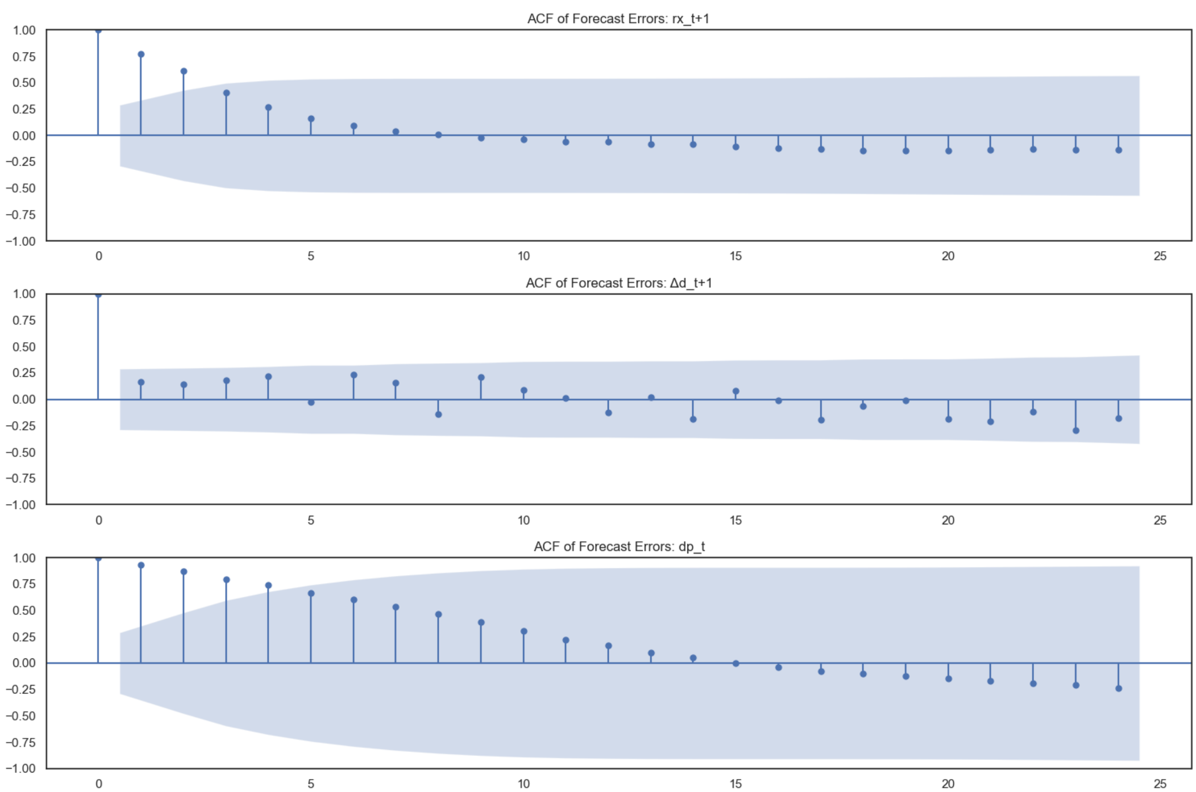


Figure 13

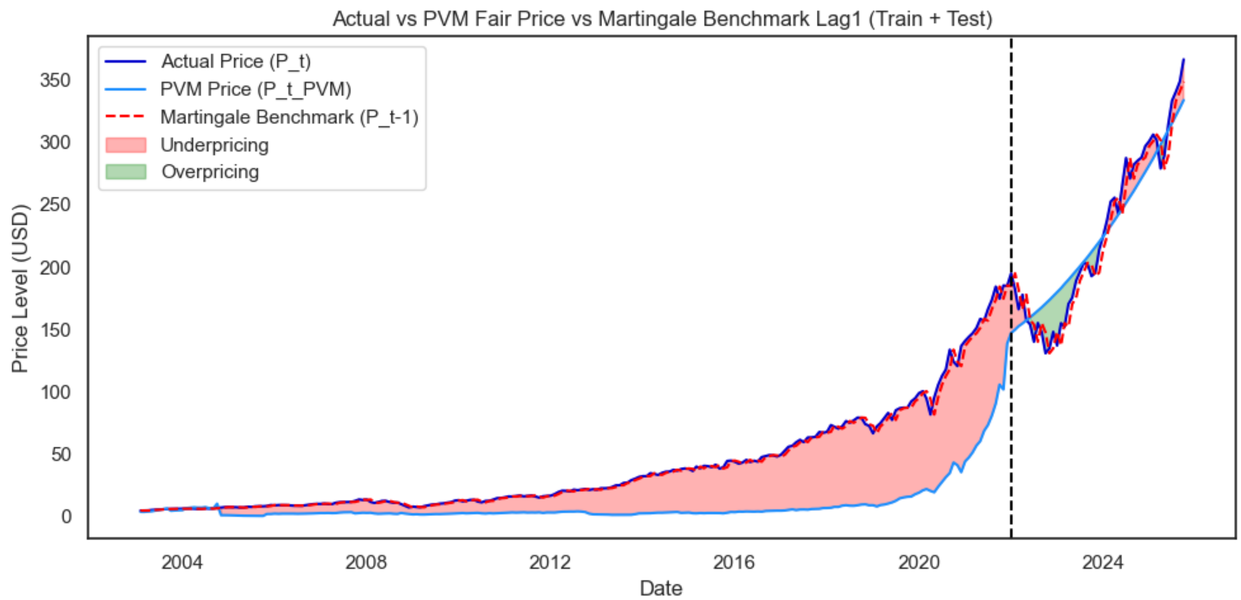


Figure 14

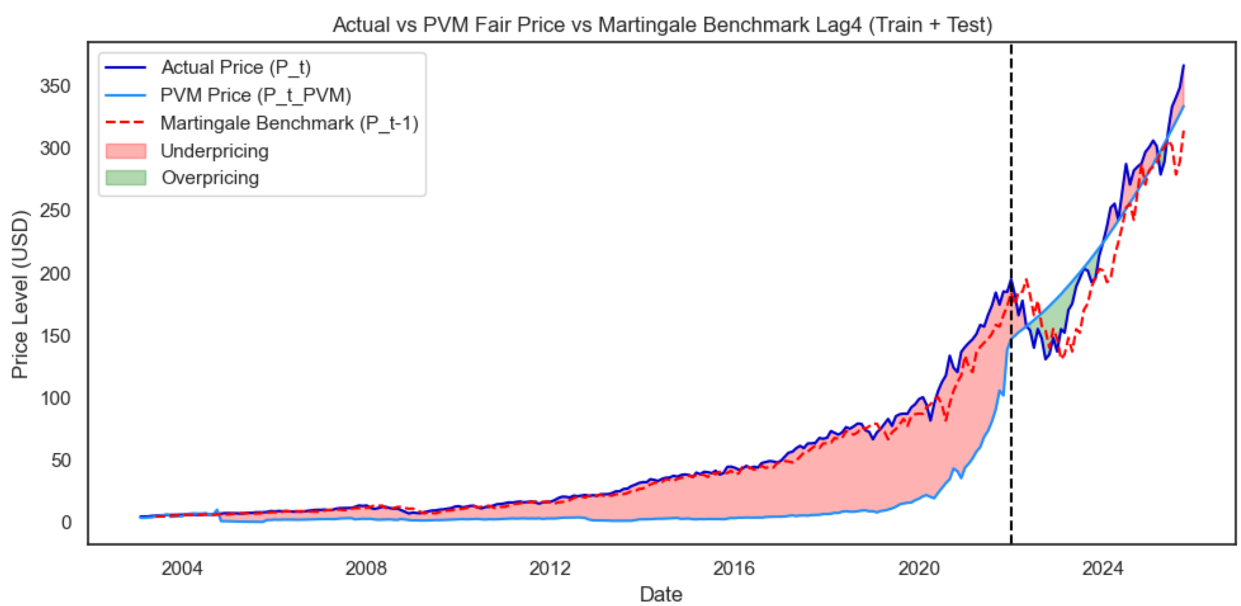


Figure 15

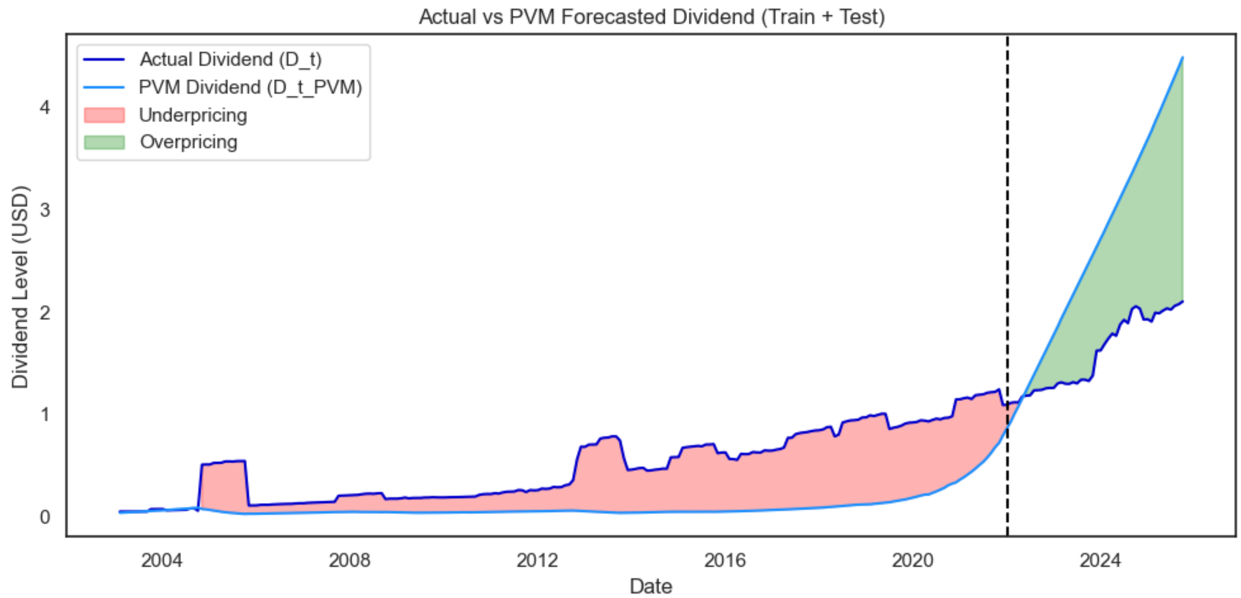


Figure 16

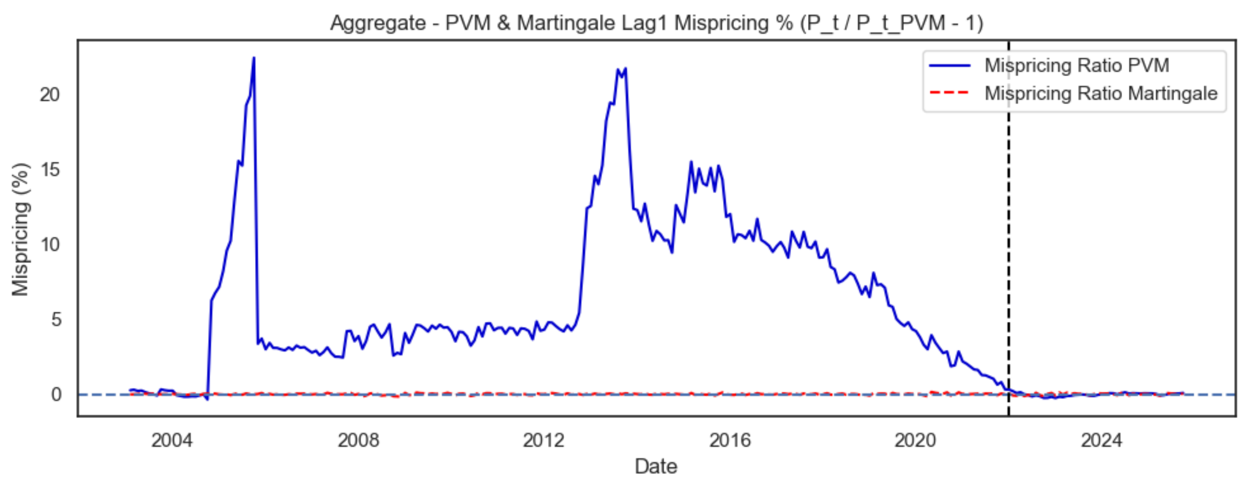


Figure 17

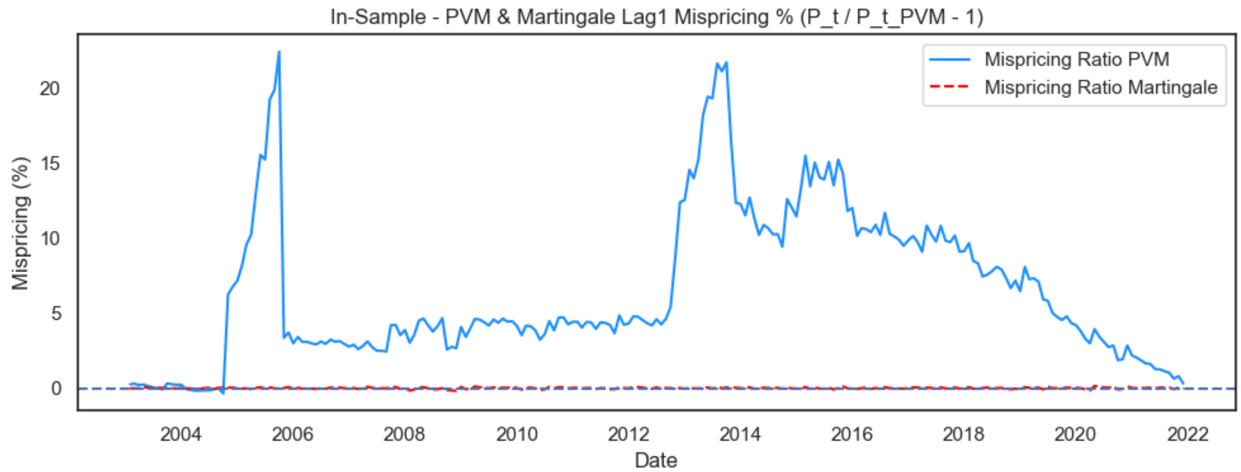


Figure 18

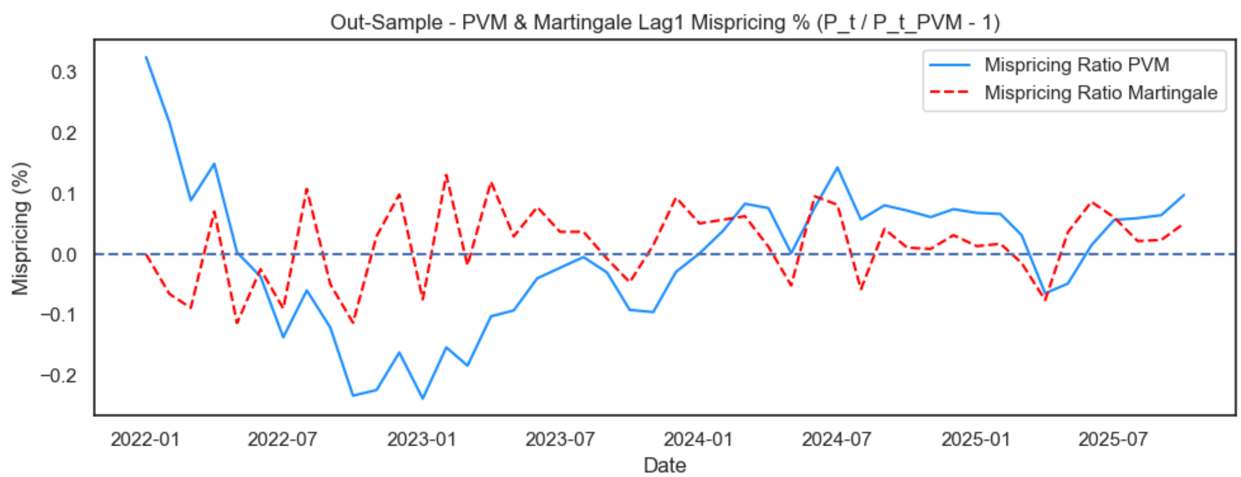


Figure 20

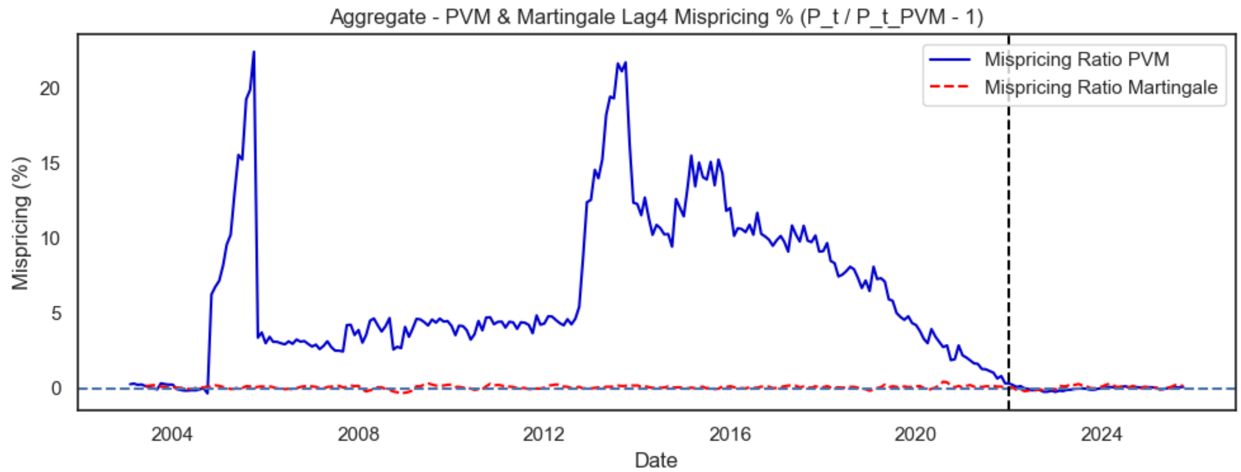


Figure 21

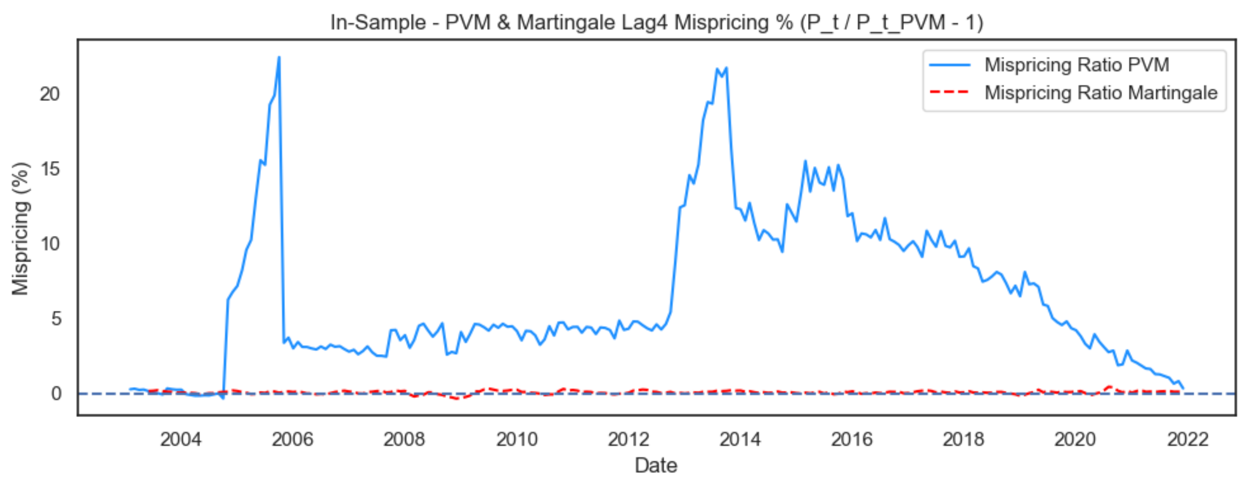
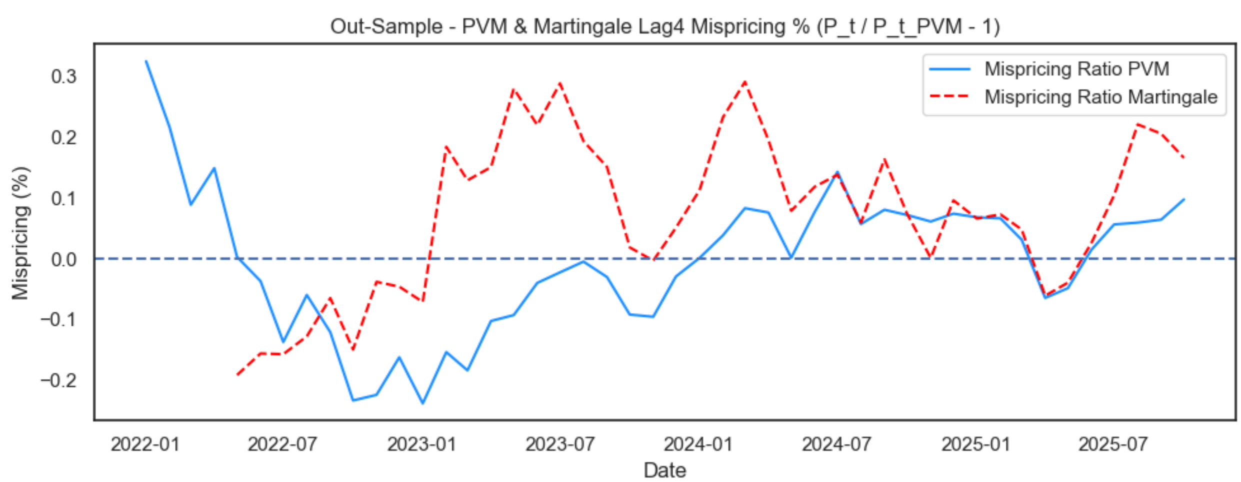


Figure 22



Proof 1

Assume that $(P_t)_{t\geq 0}$ is a martingale with respect to $(\mathcal{F}_t)$, so that

$$E[P_t \mid \mathcal{F}_{t-1}] = P_{t-1}.$$

Define the simple return by

$$r_t = \frac{P_t - P_{t-1}}{P_{t-1}}.$$

Then

$$E[r_t \mid \mathcal{F}_{t-1}] = E\left[\frac{P_t - P_{t-1}}{P_{t-1}} \,\middle|\, \mathcal{F}_{t-1}\right] = \frac{1}{P_{t-1}}\, E[P_t - P_{t-1} \mid \mathcal{F}_{t-1}]\,.$$

Since P_{t-1} is $\mathcal{F}_{t-1}$–measurable and $E[P_t \mid \mathcal{F}_{t-1}] = P_{t-1}$, we obtain

$$E[r_t \mid \mathcal{F}_{t-1}] = \frac{1}{P_{t-1}}\,(P_{t-1} - P_{t-1}) = 0.$$

Thus

$$E[r_t \mid \mathcal{F}_{t-1}] = 0, \qquad \forall t.$$

The same argument applies to log returns, since

$$\log r_t = \log P_t - \log P_{t-1},$$

and applying conditional expectation together with the martingale property yields a conditional mean of zero.

Proof 2

(a) Each S_t^i is a martingale, and the process

$$I_t = \sum_{i=1}^{n} w_i S_t^i$$

is simply a weighted sum of them. Since the weights w_i are constants, I_t is adapted. For any $s \le t$,

$$E[I_t \mid \mathcal{F}_s] = E\left[\sum_{i=1}^{n} w_i S_t^i \,\middle|\, \mathcal{F}_s\right] = \sum_{i=1}^{n} w_i \, E[S_t^i \mid \mathcal{F}_s] = \sum_{i=1}^{n} w_i S_s^i = I_s.$$

Thus (I_t) is a martingale.

(b) The argument is identical to part (a), except the sum is taken only over $i \in J$. Therefore,

$$E[J_t \mid \mathcal{F}_s] = \sum_{i \in J} a_i \, E[S_t^i \mid \mathcal{F}_s] = \sum_{i \in J} a_i S_s^i = J_s,$$

so (J_t) is a martingale.

(c) We are given that each

$$X_t^i = S_t^i + D_t^i$$

is itself a martingale. Define

$$K_t = \sum_{i \in J} b_i X_t^i.$$

Since this is again a linear combination with constant coefficients,

$$E[K_t \mid \mathcal{F}_s] = \sum_{i \in J} b_i \, E[X_t^i \mid \mathcal{F}_s] = \sum_{i \in J} b_i X_s^i = K_s.$$

Hence (K_t) is a martingale.