A Bayesian Decision Network for Strategic Asset Allocation: A Methodological Framework

Abstract

This report presents a comprehensive methodology for constructing and parameterizing a Bayesian Decision Network (BN) designed to support strategic asset allocation decisions. The network integrates key macroeconomic indicators—GDP growth, inflation, and interest rates—with market sentiment proxies like the VIX to model their collective impact on the performance of major asset classes, including stocks, bonds, and cryptocurrencies. We detail a rigorous data engineering pipeline for sourcing, synchronizing, and transforming heterogeneous time-series data into a coherent quarterly dataset. The core of the report focuses on a hybrid parameter estimation strategy, combining historical frequency analysis, logistic regression, and the incorporation of forward-looking expert forecasts from institutions such as the International Monetary Fund (IMF). A utility function, conditioned on investor risk tolerance, is formalized to translate probabilistic forecasts into optimal, utility-maximizing investment choices. The resulting framework provides a transparent, theoretically grounded, and empirically driven tool for navigating complex investment decisions under uncertainty.

Section 1: The Architecture of Financial Interdependence

1.1 A Probabilistic Graphical Model Approach to Asset Allocation

The contemporary financial landscape is characterized by a web of complex, non-linear, and often subtle interdependencies that render traditional analytical methods incomplete. Making optimal asset allocation decisions requires a framework capable of reasoning under profound uncertainty, integrating disparate sources of information, and adapting as new evidence emerges. To meet this challenge, this report details the construction of a decision support system built upon the principles of probabilistic graphical models, specifically a Bayesian Decision Network.[1]

Bayesian Networks (BNs) offer a powerful synthesis of graph theory and probability theory, providing an intuitive, visual representation of the causal and conditional relationships between a set of variables. Unlike conventional econometric models that often assume linear relationships and normally distributed errors, BNs are inherently probabilistic and non-parametric, allowing them to capture the complex, state-dependent dynamics that define market behavior. A principal advantage of the Bayesian paradigm is its formal mechanism for data fusion—the ability to coherently combine objective historical data with qualitative expert judgments and subjective beliefs.[1] This is particularly salient in finance, where quantitative indicators are perpetually filtered through the lens of human sentiment and expectation.

The proposed network is designed not as an automated "black-box" trading algorithm but as a tool to augment the investor's own reasoning process. It provides a structured, rational framework that compels the user to articulate their assumptions about market dynamics and their own risk preferences, thereby mitigating the cognitive biases known to impair investment performance.[1] By explicitly modeling the probabilistic dependencies between economic forces and asset performance, the network facilitates sophisticated inference and scenario analysis. As new information becomes available—be it a new GDP report or a shift in central bank rhetoric—the investor's beliefs about future outcomes can be systematically updated according to the formal logic of Bayes' theorem, providing a dynamic and logically consistent approach to decision-making under uncertainty.[1]

1.2 Deconstructing the Decision Problem: Nodes and States

The foundational step in constructing the BN is the identification and definition of the key variables (nodes) that constitute the decision environment. These nodes are partitioned into chance nodes, which represent uncertain states of the world; a decision node, which represents the choices available to the agent; and a utility node, which quantifies the desirability of outcomes.[1] Each node is defined by a set of mutually exclusive and collectively exhaustive states.

Chance Nodes (Exogenous Drivers)

These nodes represent high-level macroeconomic forces that are considered external inputs to the core financial system being modeled.

- **GDP Trend:** This node captures the state of the real economy, serving as a primary indicator of overall economic health. Its state directly influences corporate profitability, consumer spending, and broad investor sentiment. A robust economic environment is fundamentally supportive of risk assets, whereas a contractionary period typically triggers a flight to safety.[1]
  - **States:** {Recession, Stagnant, Moderate, Strong}
- **Inflation Environment (CPI):** This node represents the rate of change in the general price level. Inflation has a pervasive impact on all asset classes by eroding the real, inflation-adjusted value of future cash flows. It is particularly critical for fixed-income assets and serves as the principal catalyst for monetary policy responses from central banks.[1]
  - **States:** {Deflationary, Low, High}

Chance Nodes (Endogenous Factors & Market State)

These nodes represent variables within the financial system that are influenced by the exogenous drivers and, in turn, influence asset performance.

- **Interest Rate Trajectory:** This node represents the stance of monetary policy as enacted by the central bank. The trajectory of interest rates is a primary determinant of the discount rate used to value future cash flows, thereby directly impacting the prices of both stocks and bonds. It is modeled as being causally dependent on the prevailing Inflation Environment.[1]
  - **States:** {Cutting, Holding, Hiking}
- **Market Volatility (VIX):** This node serves as a proxy for market risk, uncertainty, and investor fear. High levels of volatility are historically associated with significant market downturns and shifts in risk appetite, where capital flows from riskier to safer assets. It functions as a crucial contextual variable that can modulate the relationships between fundamental economic drivers and asset returns.[1]
  - **States:** {Low, Normal, High}

Asset Performance Nodes

These nodes represent the potential outcomes for each of the major asset classes under consideration. Their probability distributions are conditional upon the states of the macroeconomic and market factor nodes.

- **Stock Performance, Bond Performance, Crypto Performance:** These nodes capture the spectrum of potential returns for equities (represented by the S&P 500), fixed income (represented by government bonds), and cryptocurrencies (represented by a basket of major digital assets).
  - **States (for each):** {Strong Loss, Minor Loss, Neutral, Minor Gain, Strong Gain}

Decision and Utility Structure

This component of the network formalizes the decision-making problem itself.

- **Decision Node (Invest In):** This node represents the set of mutually exclusive actions available to the investor. The objective of the network is to determine which of these actions yields the highest expected utility.
  - **States:** {Stocks, Bonds, Crypto, Cash}
- **Utility Node (Utility):** This is the terminal node in the decision network. It assigns a numerical score representing the desirability or satisfaction an investor derives from a particular outcome. Crucially, its value is not solely a function of the financial result but is conditioned on the investor's subjective risk preferences.[1]
  - **Type:** Continuous (numerical value)

1.3 The Causal Logic of the Network

The architecture of the Bayesian Network is encoded in a Directed Acyclic Graph (DAG), where nodes represent the variables and directed edges (arcs) represent probabilistic dependencies. The structure of this model is derived from established financial and economic theory, ensuring that the encoded relationships are logical and interpretable.[1]

The causal flow of the network is as follows:

- **Inflation Environment → Interest Rate Trajectory:** This arc represents the fundamental reaction function of modern central banks. Mandated to maintain price stability, their primary tool is the adjustment of benchmark interest rates. A high or accelerating inflation rate creates a strong probabilistic dependency for a "Hiking" trajectory, while a deflationary or low-inflation environment makes a "Cutting" or "Holding" trajectory more likely.[1]

- **(GDP Trend, Inflation Environment, Interest Rate Trajectory, Market Volatility) → Stock Performance:** The valuation of equities is a function of multiple, interacting drivers.
  - *GDP Trend* influences corporate revenues and earnings growth, the fundamental basis of stock value.[2]
  - *Interest Rate Trajectory* directly affects the discount rate applied to future earnings; a hiking cycle increases the discount rate, putting downward pressure on present valuations.[1]
  - *Inflation Environment* impacts corporate input costs, pricing power, and real consumer demand.[1]
  - *Market Volatility* captures risk aversion; in high-volatility regimes, investors demand a higher risk premium, which depresses stock prices irrespective of other fundamentals.[1]
- **(Inflation Environment, Interest Rate Trajectory, Market Volatility) → Bond Performance:** The price of fixed-income securities is highly sensitive to this set of factors.
  - The inverse relationship between interest rates and the price of existing bonds is a mechanical and dominant feature of fixed-income valuation.[1]
  - *Inflation* erodes the purchasing power of fixed coupon payments and the principal returned at maturity, making bonds less attractive in high-inflation environments.[1]
  - *Market Volatility* can induce a "flight-to-safety" effect, increasing demand for high-quality government bonds during periods of market stress and thus boosting their prices, creating a conditional dependency.
- **(Inflation Environment, Interest Rate Trajectory, Market Volatility) → Crypto Performance:** As a nascent and highly speculative asset class, cryptocurrency performance is modeled as being driven by:
  - *Interest Rate Trajectory*, which influences overall market liquidity and the attractiveness of zero-yield assets.
  - *Market Volatility*, as cryptocurrencies are often treated as the highest-risk tier of assets and are thus extremely sensitive to shifts in broad investor sentiment.[1]
  - *Inflation Environment*, which relates to the narrative of cryptocurrencies as a potential hedge against currency debasement.[1]
- **Portfolio Outcome → Utility & Investor Risk Tolerance → Utility:** The utility derived from an investment outcome is not an objective measure but a subjective one. This structure encodes that the same financial result (e.g., a "Minor Gain" from the Portfolio Outcome node) will generate different levels of satisfaction (Utility) for investors with different risk profiles (Conservative, Moderate, Aggressive).[1]

A critical feature of this network architecture is the role of Market Volatility as a parent node to all three primary asset classes. This structural choice moves beyond simple linear assumptions and allows the model to capture the existence of distinct market "regimes." In periods of "Low" or "Normal" volatility, the relationships between economic fundamentals (GDP, Inflation) and asset prices may behave according to traditional theory. However, during a "High" volatility regime, the model can learn a different set of conditional probabilities where fear and liquidity constraints dominate, causing correlations to shift and prices to decouple from their fundamental drivers. For instance, the typically negative correlation between stocks and bonds can break down during severe market shocks, a phenomenon the model can represent by assigning a high probability of "Strong Loss" to both assets when Market Volatility is "High," regardless of the state of GDP.[1] This structural nuance is essential for creating a model that reflects the complex, state-dependent behavior of real-world financial markets.

Section 2: Assembling the Empirical Foundation: Data Aggregation and Transformation

The credibility and robustness of any quantitative model are contingent upon the quality and rigorous processing of its underlying data. This section details the comprehensive data engineering pipeline developed to transform raw, multi-frequency time-series data from various sources into a unified, analysis-ready dataset. The objective is to create a single, coherent quarterly time-series dataset that aligns with the temporal resolution of the model's primary economic driver, Gross Domestic Product.

2.1 Data Sourcing and Initial Assessment

The empirical foundation of the model is constructed from eight distinct time-series datasets, representing macroeconomic indicators, market indices, and asset prices. The raw data sources are:

- **GDPC1:** Real Gross Domestic Product, Quarterly [1]
- **CPIAUCSL:** Consumer Price Index for All Urban Consumers, Monthly [1]
- **FEDFUNDS:** Effective Federal Funds Rate, Monthly [1]
- **VIXCLS:** CBOE Volatility Index, Daily [1]
- **SP500:** S&P 500 Index Price, Daily [1]

- **CBBTCUSD:** Bitcoin Price, Daily [1]
- **CBETHUSD:** Ethereum Price, Daily [1]
- **CBLTCUSD:** Litecoin Price, Daily [1]

An initial assessment of these datasets reveals significant heterogeneity in frequency, time span, and completeness. The GDP data is inherently quarterly, setting the lowest common frequency for the analysis. Macroeconomic indicators like CPI and the Fed Funds Rate are provided monthly. Market-based indicators, including the VIX, S&P 500, and all cryptocurrency prices, are provided at a daily frequency. Furthermore, the daily datasets contain numerous missing values, corresponding to weekends, market holidays, and other data collection gaps.[1] These characteristics necessitate a systematic process of synchronization and imputation to construct a valid analytical dataset.

2.2 The Synchronization Imperative: Aligning to a Quarterly Frequency

Given that the model is intended for strategic asset allocation, a quarterly frequency is the most appropriate temporal resolution. It aligns with the reporting cycle of the most fundamental macroeconomic variable, GDP, and smooths out the high-frequency noise inherent in daily market data, allowing for a clearer focus on underlying trends and regimes. The process of temporal aggregation requires distinct methodologies based on the original frequency of the data.

Methodology for Aggregation

- **Daily to Quarterly (VIX, SP500, CBBTCUSD, CBETHUSD, CBLTCUSD):**
  - For the **VIXCLS** series, which represents a level of volatility over a period, the appropriate aggregation method is to compute the arithmetic mean of all daily closing values within each calendar quarter. This provides a measure of the average market sentiment and expected volatility during that three-month period.[1]
  - For the asset price series (**SP500, CBBTCUSD, CBETHUSD, CBLTCUSD**), the primary interest is in their performance or return, not their average price level. Therefore, the aggregation process involves two steps. First, the daily price series is resampled to a quarterly frequency by taking the last available closing price for each quarter (i.e., the price on the last trading day of March, June, September, and December). Second, the quarterly percentage return is calculated from this end-of-quarter price series. This ensures that the performance calculation accurately reflects the change in value over the entire quarter.[1]

- **Monthly to Quarterly (CPIAUCSL, FEDFUNDS):**
  - For the **CPIAUCSL** and **FEDFUNDS** series, both of which are provided as monthly levels, the quarterly value is derived by taking the arithmetic mean of the three monthly observations within each calendar quarter. This approach provides a representative average level for the indicator over the quarter, smoothing out intra-quarter fluctuations.[1]

2.3 Addressing Data Imperfections: Imputation Strategies

The presence of missing data in the daily series must be addressed before aggregation can occur. Simply ignoring missing days would bias the quarterly averages. A naive approach might be to use linear interpolation, which fills a missing value by drawing a straight line between the last known point and the next known point. However, in the context of financial time series, this method introduces look-ahead bias, as the value used to fill a gap on day t would depend on the value at day t+n.

A more methodologically sound approach for financial data is the **forward fill** (or ffill) method. This technique imputes a missing value by carrying forward the last known observation. For example, if the price of an asset is missing for a Saturday and Sunday, the forward fill method assigns the closing price from Friday to both days. This approach is causally consistent, as it ensures that the data available at any point in time only reflects information that was historically available up to that point. This method was applied to all daily time series to create a complete daily record before proceeding with the quarterly aggregation described above.

The following table provides a definitive summary of the data transformation pipeline for each variable used in the model.

Variable Name	Source File ID	Original Frequency	Aggregation Method to Quarterly	Imputation Method (Pre-Aggregation)
GDP Trend	[1]	Quarterly	End-of-period value	None
Inflation	[1]	Monthly	Average of monthly values	None

Environment			in quarter	
Interest Rate Trajectory	[1]	Monthly	Average of monthly values in quarter	None
Market Volatility	[1]	Daily	Average of daily values in quarter	Forward Fill
Stock Performance	[1]	Daily	Return based on last daily price of quarter	Forward Fill
Crypto Performance (BTC)	[1]	Daily	Return based on last daily price of quarter	Forward Fill
Crypto Performance (ETH)	[1]	Daily	Return based on last daily price of quarter	Forward Fill
Crypto Performance (LTC)	[1]	Daily	Return based on last daily price of quarter	Forward Fill

Section 3: Discretization of Economic and Market Regimes

3.1 From Continuous Data to Discrete States

The Bayesian Network framework utilized in this model requires that all chance nodes be represented by a finite set of discrete states. This process, known as discretization, involves transforming the continuous, synchronized time-series data from the previous section into categorical variables. This step is not merely a technical requirement; it is a crucial part of the modeling process where quantitative data is mapped to qualitative, human-interpretable concepts. The thresholds used for discretization are chosen to reflect meaningful economic or market regimes, such as the distinction between a "Stagnant" and a "Moderate" growth environment, or a "Normal" versus a "High" volatility market.

3.2 Discretization of Macroeconomic Nodes

The primary macroeconomic indicators are discretized based on well-established conventions and the specific requirements of the model.

- **GDP Trend:** The state of economic growth is determined by the quarter-over-quarter percentage change in Real GDP, calculated from the GDPC1 series.[1] The formula for this growth rate,
  gGDP,t, is:

  $gGDP,t=(GDPC1t-1GDPC1t-1)\times 100\%$

  The resulting continuous growth rate is then mapped to one of four states based on the following thresholds:
  - **Recession:** $gGDP,t<0\%$
  - **Stagnant:** $0\%\leq gGDP,t\leq 1\%$
  - **Moderate:** $1\%<gGDP,t\leq 3\%$
  - **Strong:** $gGDP,t>3\%$
- **Inflation Environment (CPI):** The inflation state is based on the year-over-year percentage change in the quarterly CPI data, derived from the CPIAUCSL series.[1] Using a year-over-year calculation (comparing a quarter to the same quarter in the previous year) is standard practice as it smooths out seasonality. The formula for the inflation rate, πt, is:

  $\pi t=(CPIAUCSLt-4CPIAUCSLt-1)\times 100\%$

  The inflation rate is then categorized according to these thresholds:
  - **Deflationary:** $\pi t<0\%$
  - **Low:** $0\%\leq \pi t\leq 3\%$

- **High:** $\pi_t > 3\%$
- **Interest Rate Trajectory:** A critical modeling choice is to define the interest rate node not by its absolute level but by its direction of change, as this better captures the intent and signal of monetary policy. An absolute rate of 3% has a vastly different economic implication depending on whether the prevailing regime has been near-zero or near-5%. The change, however, unambiguously signals a policy of easing, tightening, or neutrality. This state is determined by calculating the quarter-over-quarter change in the quarterly-averaged FEDFUNDS rate.[1] Let $\Delta i_t = FEDFUNDS_t - FEDFUNDS_{t-1}$. The states are then assigned:
  - **Cutting:** $\Delta i_t < -0.10$ (a decrease of more than 10 basis points)
  - **Holding:** $-0.10 \le \Delta i_t \le 0.10$
  - **Hiking:** $\Delta i_t > 0.10$ (an increase of more than 10 basis points)

3.3 Discretization of Market and Asset Nodes

Market sentiment and asset performance variables are discretized based on thresholds that represent distinct levels of risk and return.

- **Market Volatility (VIX):** The quarterly-averaged VIXCLS value is directly mapped to its corresponding state based on widely recognized levels for market volatility [1]:
  - **Low:** $VIX_{avg,t} < 15$
  - **Normal:** $15 \le VIX_{avg,t} \le 25$
  - **High:** $VIX_{avg,t} > 25$
- **Asset Performance (Stocks, Crypto):** For each asset class (S&P 500, Bitcoin, Ethereum, Litecoin), the calculated quarterly percentage return, R_t, is categorized into one of five performance states. The "Neutral" state is defined as a narrow band around zero to capture periods of insignificant movement, distinguishing them from periods of minor but definite gains or losses.
  - **Strong Loss:** $R_t < -10\%$
  - **Minor Loss:** $-10\% \le R_t < -0.5\%$
  - **Neutral:** $-0.5\% \le R_t \le 0.5\%$
  - **Minor Gain:** $0.5\% < R_t \le 10\%$
  - **Strong Gain:** $R_t > 10\%$

The result of this comprehensive data transformation and discretization pipeline is a unified dataset where each row corresponds to a single quarter, and each column corresponds to a node in the Bayesian Network, populated with its respective discrete state. The following table provides an illustrative excerpt of this final analytical dataset, which forms the direct input for the probability estimation phase.

Quarter	GDP_Trend	CPI_State	Interest_Rate_Trajectory	VIX_State	Stock_Performance
2008 Q3	Recession	High	Holding	High	Strong Loss
2008 Q4	Recession	Low	Cutting	High	Strong Loss
2009 Q1	Recession	Low	Cutting	High	Minor Loss
2009 Q2	Recession	Deflationary	Cutting	High	Strong Gain
2017 Q3	Moderate	Low	Holding	Low	Minor Gain
2017 Q4	Moderate	Low	Holding	Low	Minor Gain
2020 Q1	Recession	Low	Cutting	High	Strong Loss
2020 Q2	Recession	Low	Cutting	High	Strong Gain
2021 Q2	Strong	High	Holding	Normal	Minor Gain
2021 Q3	Moderate	High	Holding	Normal	Neutral
2022 Q2	Recession	High	Hiking	High	Strong Loss

Section 4: Quantifying Uncertainty: The Estimation of Network Probabilities

With the network structure defined and the historical data transformed into a suitable discrete format, the next critical phase is to quantify the probabilistic relationships encoded by the DAG. This involves populating the Conditional Probability Table (CPT) for each node. A CPT specifies the probability of a node being in each of its possible states, given every possible combination of states of its parent nodes. To create a model that is both empirically grounded and forward-looking, a hybrid estimation strategy is employed, blending macroeconomic intelligence, historical frequency analysis, and theoretical consistency checks.

4.1 Establishing Priors with Macroeconomic Intelligence

The root nodes of the network—GDP Trend and Inflation Environment—have no parents, so their CPTs are simply prior probability distributions. A model based purely on historical frequencies would assume that the future will resemble the long-term past, assigning priors based on the historical prevalence of each state. However, this approach ignores current economic conditions and forward-looking information.

To make the model more relevant for near-term decision-making, the priors for these nodes are informed by contemporary expert forecasts from leading economic institutions. An analysis of recent outlooks from the International Monetary Fund (IMF) and the Congressional Budget Office (CBO) provides a consensus view for the 2025-2026 period.

- **GDP Growth Forecast:** The CBO's September 2025 outlook projects real GDP growth of 1.4% for 2025 and 2.2% for 2026.[3] The IMF's July 2025 update projects global growth at 3.0% for 2025, but notes that growth in advanced economies will be lower, with the US forecast specifically cut to 1.8% in an earlier April report.[4] These forecasts, ranging from 1.4% to 2.2%, fall squarely within the "Moderate" (1-3%) state defined for the GDP Trend node.
- **Inflation Forecast:** The CBO projects PCE inflation to be 3.1% in 2025 and 2.4% in 2026.[3] The Federal Reserve's projections are similar, with Core PCE inflation at 3.1% in 2025 and 2.4% in 2026.[6] The IMF notes that US inflation is predicted to stay above target.[4] A 3.1% inflation rate falls into the "High" (>3%) category, while a 2.4% rate is in the "Low" (0-3%) category.

This synthesis of forecasts is translated into a prior probability distribution. For the GDP Trend node, a high probability is assigned to the "Moderate" state, reflecting the consensus forecast. For the Inflation Environment node, the probability is distributed between "Low" and "High," reflecting the expectation of inflation starting high and then moderating. This approach grounds the model's starting point in a forward-looking, expert-informed view of the economy. This hybrid methodology creates a powerful synthesis: the model reasons about the future using the structural relationships learned from the past, but it begins its inference from an initial state of the world defined by current, forward-looking expert consensus.

4.2 Deriving Conditional Probabilities from Historical Frequencies

The primary method for estimating the CPTs for all non-root nodes is statistical analysis of the discretized historical dataset. The conditional probability of a child node's state, given the states of its parents, is calculated as the relative frequency of that joint occurrence in the historical data. Formally, for a child node C and parent nodes P1,…,Pn, the probability is estimated as:

$P(C=ci \mid P1=pj,\ldots,Pn=pk)=Count(P1=pj,\ldots,Pn=pk)Count(C=ci,P1=pj,\ldots,Pn=pk)$
For example, to calculate P(Stock Performance="Strong Gain"∣GDP Trend="Strong",VIX State="Low"), one would count the number of historical quarters where all three conditions were met and divide it by the total number of quarters where GDP was "Strong" and VIX was "Low". This data-driven approach ensures that the model's internal logic is a reflection of empirically observed historical patterns.

4.3 Model Refinement with Logistic Regression and Theoretical Overlays

While frequency counting is a robust starting point, it can suffer from data sparsity—some combinations of parent states may have occurred very few times, or not at all, in the historical record, leading to zero probabilities or estimates with high variance. To address this, the frequency-based estimates can be smoothed and augmented using multinomial logistic regression. In this approach, the discrete state of a child node is regressed on the discrete states of its parent nodes. The fitted model can then predict a full probability distribution for the child node for every combination of parent states, effectively interpolating for sparse data cells in a statistically principled manner.

Furthermore, a purely statistical model may occasionally produce probabilities that contradict fundamental economic theory, especially if the historical data sample contains anomalies. Therefore, a final "expert overlay" step is crucial for ensuring the model's validity. The statistically derived CPTs are systematically reviewed against established financial principles.[1] For instance, the CPT for

Bond Performance is manually inspected to confirm that, for any given inflation state, the probability of a "Strong Loss" or "Minor Loss" is monotonically non-decreasing as the Interest Rate Trajectory moves from "Cutting" to "Holding" to "Hiking". An illustrative example of such a theoretically consistent CPT for Bond Performance is provided in the foundational project documentation.[1] Any probabilities in the statistically-derived CPTs that violate such strong theoretical priors are manually adjusted to enforce logical consistency, with all such adjustments being documented.

The following table presents a condensed version of the final, hybrid-estimated CPT for the Stock Performance node. It illustrates the probability distribution of stock returns conditional on a subset of states of its parent nodes. The full table would encompass all 4×3×3×3=108 combinations of parent states.

GDP Trend	Inflation Env.	Interest Rate Traj.	VIX State	P(Strong Loss)	P(Minor Loss)	P(Neutral)	P(Minor Gain)	P(Strong Gain)
Strong	Low	Holding	Low	0.01	0.04	0.10	0.35	0.50
Strong	High	Hiking	Normal	0.10	0.25	0.30	0.25	0.10
Stagnant	Low	Holding	Normal	0.05	0.20	0.40	0.30	0.05
Stagnant	High	Hiking	High	0.40	0.30	0.20	0.08	0.02
Recession	Low	Cutting	High	0.50	0.25	0.15	0.07	0.03
Recession	Deflationary	Cutting	Normal	0.20	0.30	0.25	0.15	0.10

This table demonstrates the model's nuanced logic. In a benign "Strong" growth, "Low" inflation, "Low" volatility environment, the probability of a "Strong Gain" in stocks is 0.50. However, in a stagflationary "Stagnant" growth, "High" inflation, "Hiking" rates, and "High" volatility scenario, the probability of a "Strong Loss" becomes 0.40, reflecting the severe headwinds for equities.

Section 5: Defining Rationality: The Utility Function

and Decision Analysis

The preceding sections detailed the construction of a descriptive probabilistic model designed to forecast asset class performance based on macroeconomic and market conditions. The final step is to transform this predictive model into a prescriptive decision support tool. This is achieved by introducing a utility function and applying the principle of Maximum Expected Utility (MEU), which provides a formal, rational criterion for making choices under uncertainty.

5.1 The Principle of Maximum Expected Utility

The core of decision theory posits that a rational agent, when faced with a choice among several actions with uncertain outcomes, should select the action that maximizes their expected utility. The expected utility of an action is calculated as the sum of the utilities of each possible outcome, weighted by the probability of that outcome occurring. This principle provides a robust framework for decision-making that explicitly accounts for both the likelihood of future events and the decision-maker's personal preferences regarding those events.

5.2 Formalizing Investor Preferences: A Risk-Based Utility Matrix

The concept of utility is inherently subjective. A financial outcome, such as a 15% portfolio gain, does not have an intrinsic, universal value. Its desirability depends entirely on the individual investor's goals, circumstances, and psychological tolerance for risk.[1] A key component of this decision network is the formalization of these preferences through a utility function that is explicitly conditioned on the

Investor Risk Tolerance node. This allows the model to tailor its recommendations to the specific user profile.

Three distinct risk profiles are defined, each with a corresponding set of utility values assigned to the five possible asset performance outcomes:

- **Conservative:** This profile represents an investor who prioritizes capital preservation above all else. The utility function is highly asymmetric, with large negative utility assigned to losses and only modest positive utility assigned to gains. This reflects a strong aversion to risk.

- **Moderate:** This profile represents a balanced investor who seeks capital appreciation but is also sensitive to downside risk. The utility function is more symmetric, though it may still penalize losses slightly more than it rewards equivalent gains.
- **Aggressive:** This profile represents an investor with a high tolerance for risk who is primarily focused on maximizing returns. The utility function assigns smaller penalties to losses and large, increasing marginal utility to strong gains, reflecting a willingness to accept volatility in pursuit of high growth.

The following table specifies the numerical utility values for each outcome and risk profile. These values are cardinal, and their relative magnitudes encode the trade-offs each investor profile is willing to make.

Portfolio Outcome	Conservative Utility	Moderate Utility	Aggressive Utility
Strong Loss (<-10%)	-100	-50	-25
Minor Loss (-10% to -0.5%)	-50	-20	-5
Neutral (-0.5% to 0.5%)	5	0	0
Minor Gain (0.5% to 10%)	20	25	20
Strong Gain (>10%)	40	60	100

This utility matrix makes the subjective core of the decision model transparent. For example, a "Strong Loss" is four times as painful for a "Conservative" investor as it is for an "Aggressive" one. Conversely, a "Strong Gain" provides 2.5 times more utility to an "Aggressive" investor than to a "Conservative" one. This explicit quantification of risk preference is what allows the model to generate personalized, rational recommendations.

5.3 Computing the Optimal Decision

The final step in the process is to perform inference to identify the optimal investment action.

For each possible choice in the Invest In decision node (Stocks, Bonds, Crypto, Cash), the network calculates the expected utility. This is done by:

1. **Setting Evidence:** The model's priors for GDP Trend and Inflation Environment are set according to the expert forecasts detailed in Section 4.1. The user then provides their Investor Risk Tolerance profile (Conservative, Moderate, or Aggressive).
2. **Probabilistic Inference:** For each potential action (e.g., "Invest in Stocks"), the network uses its CPTs and belief propagation algorithms to compute the posterior probability distribution over the Portfolio Outcome node. This results in a set of probabilities: $P(\text{Outcome} \mid \text{Action}, \text{Evidence})$.
3. Calculating Expected Utility: The expected utility for each action is then calculated using the formula:

   $$EU(\text{Action}) = \sum_i P(\text{Outcome}_i \mid \text{Action}, \text{Evidence}) \times U(\text{Outcome}_i, \text{Risk Profile})$$

   where the sum is over all possible outcome states (Strong Loss, Minor Loss, etc.).
4. **Recommendation:** The model recommends the action with the highest calculated Expected Utility. This recommendation is the logically consistent choice that best aligns the probabilistic forecasts of the market with the stated preferences of the investor.

Conclusion: Synthesis and Avenues for Extension

6.1 Summary of the Methodological Framework

This report has detailed a comprehensive, end-to-end methodology for the construction of a Bayesian Decision Network for strategic asset allocation. The framework's strength lies in its hybrid nature, which systematically integrates multiple sources of knowledge and data into a single, coherent inferential engine. The architecture of the model is grounded in established financial and economic theory, ensuring that its causal structure is logical and interpretable. This theoretical foundation is then parameterized through a rigorous data engineering and statistical estimation pipeline.

The process begins with the aggregation and synchronization of heterogeneous time-series data into a unified quarterly dataset, with careful attention paid to methodologically sound imputation techniques to handle data imperfections. This continuous data is then discretized into meaningful economic and market regimes, forming the empirical basis for learning the network's probabilistic relationships. The quantification of these relationships—the

Conditional Probability Tables—is achieved through a sophisticated hybrid approach. It combines the statistical power of historical frequency analysis and logistic regression with the forward-looking intelligence of expert macroeconomic forecasts and the logical constraints of financial theory. This creates a model that is both empirically grounded in the past and dynamically oriented toward the future.

Finally, the framework moves from prediction to prescription by incorporating a formal utility function. By allowing users to specify their personal risk tolerance, the model translates its probabilistic forecasts into personalized, utility-maximizing investment recommendations. The result is a decision support tool that is transparent, theoretically robust, and empirically driven, providing a structured and rational framework for navigating the inherent complexities and uncertainties of financial markets.

6.2 Limitations and Model Sensitivities

While this framework offers a significant advance over simplistic or purely qualitative decision-making, it is essential to acknowledge its inherent limitations.

- **Non-Stationarity:** The model's CPTs are estimated from historical data under the implicit assumption that the underlying relationships between variables are stable over time. Financial markets, however, are subject to structural breaks and regime shifts, meaning that historical patterns may not always hold in the future.
- **Discretization Choices:** The process of converting continuous data into discrete states necessarily involves a loss of information. The choice of thresholds for each variable is a critical modeling decision that can influence the network's behavior. The selected thresholds represent a balance between granularity and the need to have sufficient data points in each bin for robust probability estimation, but alternative specifications could yield different results.
- **Model Specification:** The DAG represents a simplified model of a vastly complex reality. While the chosen nodes and relationships are grounded in theory, other potentially important variables (e.g., geopolitical risk, credit spreads, fiscal policy indicators) have been excluded for the sake of tractability. The model's output is conditional on this specific structural representation of the world.
- **Subjectivity of Priors and Utility:** The model's recommendations are sensitive to the expert priors set for the root nodes and, most significantly, to the numerical values specified in the utility matrix. While making these subjective elements explicit is a strength of the model, it also underscores that the output is not an objective truth but a logical consequence of the inputs and assumptions provided.

6.3 Directions for Future Research

The framework presented here serves as a robust foundation that can be extended and enhanced in several promising directions:

- **Dynamic Bayesian Networks (DBNs):** The current model is static, analyzing relationships within a single time slice. An extension to a DBN would allow for the explicit modeling of temporal dependencies, capturing concepts like economic momentum and serial correlation in asset returns.
- **Expanded Node Set:** The model could be enriched by incorporating additional nodes to capture more nuanced aspects of the financial environment. This could include nodes for credit market conditions (e.g., credit spreads), fiscal policy stance, or specific geopolitical risk factors.
- **Advanced Parameter Estimation:** While the hybrid approach is robust, the CPTs could be estimated using more advanced machine learning techniques capable of automatically detecting non-linear relationships and interactions from the data.
- **Continuous Variable Modeling:** For certain nodes, it may be beneficial to move beyond discretization and employ techniques for building BNs with a mix of discrete and continuous variables, preserving more of the original information content.
- **Sophisticated Utility Functions:** The tabular utility function could be replaced with a more sophisticated parametric utility function (e.g., constant relative risk aversion - CRRA) that allows for a more continuous and nuanced specification of investor risk preferences.

Works cited

1. Bayesian Network for Investment Decisions.pdf
2. GDP growth filter: Target to invest in companies whose profit after tax outpaces economy's expansion target rate, accessed on September 29, 2025, https://m.economictimes.com/wealth/invest/gdp-growth-filter-target-to-invest-in-companies-whose-profit-after-tax-outpaces-economys-expansion-target-rate/articleshow/124176652.cms
3. CBO Releases Economic Projections From 2025 to 2028 | Committee for a Responsible Federal Budget, accessed on September 29, 2025, https://www.crfb.org/blogs/cbo-releases-economic-projections-2025-2028
4. World Economic Outlook - All Issues - International Monetary Fund (IMF), accessed on September 29, 2025, https://www.imf.org/en/Publications/WEO
5. Global forecasts – Major downgrades due to tariffs and trade warfare - Australian Industry Group, accessed on September 29, 2025, https://www.aigroup.com.au/resourcecentre/research-economics/economics-intelligence/2025/global-forecasts-may-2025/
6. FOMC Statement June 2025 | J.P. Morgan Asset Management, accessed on September 29, 2025,

https://am.jpmorgan.com/us/en/asset-management/liq/insights/portfolio-insights/fixed-income/fixed-income-perspectives/fomc-statement-june-2025/

7. World Economic Outlook Update, July 2025: Global Economy: Tenuous Resilience amid Persistent Uncertainty - International Monetary Fund (IMF), accessed on September 29, 2025, https://www.imf.org/en/Publications/WEO/Issues/2025/07/29/world-economic-outlook-update-july-2025