Agentic Portfolio Management with Black-Litterman Framework

Interim Report For Wolfe QES

Victor Xiao, Changle Li, Zhaochen Jiang

Columbia Busienss School

Apr 2nd, 2025

Research Topic

- **Overall Goal:** Develop a multi-asset allocation framework integrating heterogeneous forecasts via an *Agent-Based Black-Litterman* (ABBL) model.
- **Key Innovation:** Each asset class (Equities, Fixed Income, FX, Commodities) is forecasted by a specialized agent, and we aggregate these views in a Bayesian manner.
- **Focus:** Improve portfolio performance by systematically capturing diverse market drivers and attaching agent-specific confidence levels.

Institutional Background

- **Asset Management Landscape:**
  - Global AUM exceeds $100 trillion.
  - Growing complexity: multi-asset mandates, stricter risk controls.
- **Challenges:**
  - Traditional portfolio optimization often assumes a single forecasting model or homogeneous views.
  - Heterogeneous market participants can generate conflicting signals.

Relevance & Target Audience

- **Who Cares:**
  - Mid-sized asset managers, pension funds, private wealth managers, high-net-worth individuals.
  - Retail investors seeking institutional-grade analytics.
  - Large institutions wanting supplemental “agent-based” forecasting tools.
- **Why It Matters:**
  - **Improved Accuracy:** Specialized agents capture unique drivers (fundamentals, momentum, yield curves, etc.).
  - **Risk Management:** A Bayesian approach naturally accounts for uncertainty across multiple asset classes.
  - **Flexibility:** Allows dynamic weighting of signals based on performance and market regime.

Literature Review: Overview

- **Black-Litterman Foundations**
  - Black & Litterman (1992); He & Litterman (1999)
  - Bayesian framework fusing equilibrium returns with subjective views.
- **Agent-Based Models in Finance**
  - Hommes (2006), LeBaron (2006)
  - Heterogeneous agents explain market dynamics and deviations from rational expectations.
- **Multi-Asset Factor Models**
  - Fama & French (1993), Ang (2014)
  - Common factor exposures across equities, bonds, etc., but typically do not incorporate multiple specialized forecasting agents.

Recent Extensions: LLMs & Agent-Based Macro

- **LLM-Powered Finance**
  - Popov & Roshka (2024): FolioLLM for ETF portfolio construction with domain-specific large language models.
  - Wu et al. (2023), Yang et al. (2023): Sentiment analysis and market prediction using generative AI models.
- **Agent-Based Macroeconomics**
  - Li, Gao, & Li (2024): EconAgent for LLM-driven macro simulations.
  - Demonstrates potential for simulating policy shocks, inflation, unemployment with adaptive agents.

Our Proposed Extension

- **Gap in Literature:**
  - Traditional Black-Litterman approaches rarely incorporate multiple specialized forecasting agents in a single Bayesian structure.
  - Agent-based models often focus on price simulation, not direct portfolio construction.
- **Our Contribution:**
  - *Agent-Based Black-Litterman* (ABBL) for **multi-asset** investing.
  - Each asset class has a *dedicated forecasting agent* with unique data sources and model emphasis.
  - A Bayesian aggregator (modified Black-Litterman) fuses these heterogeneous views, using confidence scores to weight each agent's forecasts.

Contribution to Existing Literature

- **Unified Framework:**
  - Bridging **agent-based heterogeneity** and **Bayesian asset allocation**.
  - Extends the single-model limit and view inputs of classical Black-Litterman.
- **Enhanced Tactocal Allocation and Risk Management:**
  - By attaching confidence levels to each agent, the model naturally adjusts to changing market conditions.
- **Potential Real-World Impact:**
  - Automates specialized research in a single AI-driven solution.
  - Targets mid-sized managers lacking large internal research teams.

Data & Sources

- **Equity Data:**
  - CRSP, Compustat (WRDS) for returns, fundamentals (quarterly balance sheet/income statements).
- **Fixed Income & Macro:**
  - CRSP: Treasury yields, corporate bond yields, yield curve data.
  - FRED: Macro indicators (GDP, unemployment, inflation).
- **FX:**
  - Bloomberg for real-time/historical exchange rates.
  - Central bank communications.
- **Commodities:**
  - Bloomberg, CRB, or other specialized vendor for futures/spot prices, storage costs.
- **Frequency & Alignment:**
  - Typical daily, weekly, monthly or quaterly, rolling windows for calibration.
  - Outlier detection, corporate action adjustments (equities).

Data Cleaning & Merging

- **Preprocessing Steps:**
  1. **Alignment:** Merge multiple frequencies (daily, monthly) into a consistent timeline.
  2. **Missing Values:** Impute or drop, depending on data criticality and coverage.
  3. **Outlier Detection:** Winsorize extreme returns to mitigate noise or erroneous ticks.
  4. **Rolling Window Construction:**
     - Use a rolling window (e.g., previous day's data plus longer-term indicators) to estimate coefficients dynamically.
     - Minimizes look-ahead bias by restricting only to past information.
  5. **Splicing & Frequency Conversion:**
     - For monthly macro indicators, attach the *latest available* reading to daily data.
- **Quality Checks:**
  - Manual inspection of data alignment for random dates.
  - Consistency checks (e.g., sum of daily returns vs. monthly returns).

Main Empirical Techniques in Existing Literature

- **Equities:**
  - Predictive Regressions: $R_{t+1} = \alpha + \beta X_t + \epsilon_{t+1}$.
  - Factor Models (Fama-French, momentum, size, value).
- **Fixed Income:**
  - Nelson–Siegel (yields), Cochrane–Piazzesi (bond returns).
  - Emphasis on yield curve slope, curvature factors.
- **FX/Currency:**
  - UIP + macro indicators, momentum, risk sentiment.
- **Commodities:**
  - Basis (futures–spot), momentum, seasonality:
    $r_{t+1} = \alpha + \beta(f_t - s_t) + \gamma\,\mathrm{Momentum}_t + \epsilon_{t+1}$,

Agent-Based Architecture: Overview

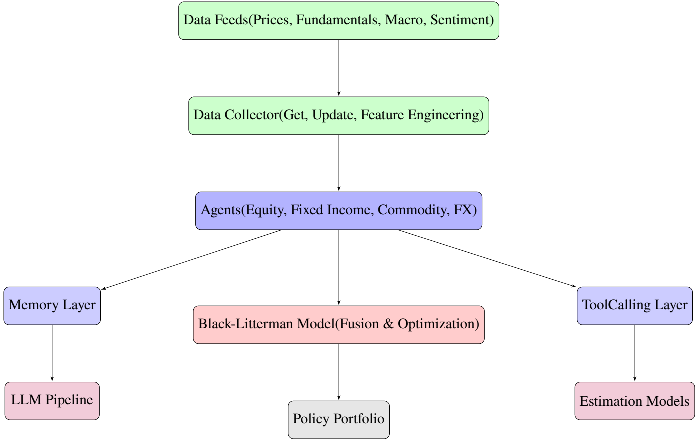


Figure 1: Agentic Portfolio Architecture

Our Proposed Empirical Setup (1/2)

- **Agent Specialization:**
  - **Equities Agent:** Combines factor signal, short-term momentum and fundamental anomalies.
  - **Bond Agent:** Uses term structure model for return prediction.
  - **FX Agent:** Blends UIP-based forecasts with interest rate differentials, macro event surprises, short-term momentum.
  - **Commodities Agent:** Emphasizes basis (futures vs. spot), roll yield, and momentum/seasonality.
- **Rolling Window Estimation:**
  - Update model parameters regularly
  - Prevents look-ahead bias, adapts to market shifts.

Our Proposed Empirical Setup (2/2)

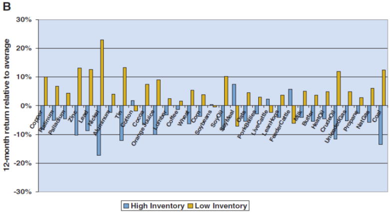
- **Confidence Scores:**
  - Each agent outputs a forecast *and* an uncertainty estimate.
  - Allows the aggregator to downweight poorly performing agents.
- **Bayesian Aggregation (Modified Black-Litterman):**
  - Each agent's forecast = separate "view."
  - Market equilibrium (often proxied by market-cap or factor-based weights) = prior distribution.
  - Combine prior + views using variance estimates to form posterior expected returns and covariance.

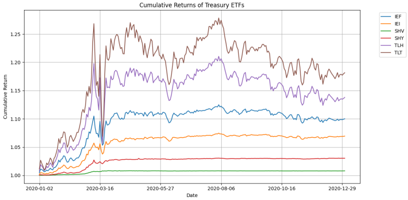
Why Our Techniques Differ

- **Heterogeneity:**
  - Instead of imposing a monotone model layer over all asset class, we allow *multiple* domain-specific signals and reasoning for interpret-ability.
  - Recognizes distinct data-generating processes (equities vs. bonds vs. FX vs. commodities).
- **Bayesian Updating:**
  - Posterior distribution adjusts to new information each period.
  - More robust to regime changes than static optimization methods.
- **End-to-end modularity and scalability:**
  - Agent utilizes domain specific data collector and feature engineering module.
  - The system is constructed with modular design in mind, allows for easy extension and implementation on any provided portfolio in the relevant asset class.

Initial Observations on Bonds

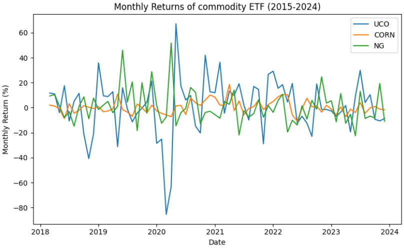
Bonds (U.S. Treasuries):

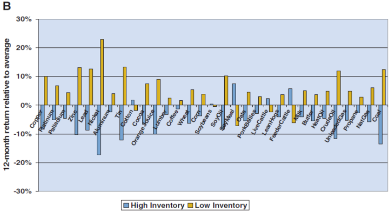




- Yield curve shifts (level, slope, curvature) correlated with macro news.
- Cochrane–Piazzesi factor shows moderate time-varying predictive power in preliminary rolling regressions.

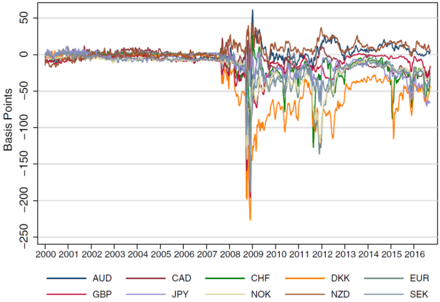
Initial Observations on Commodity





- Seasonality in agriculture and energy due to seasonal supply-demand imbalance and geopolitical events
- Basis and roll exhibits predictability with inventory level and demand.
- Commodities with *low* inventory levels tend to show higher price appreciation.

Initial Observations on FX



- Interest rate differentials matter, and evidence of momentum exists
- Short-term macro surprises drive spikes
- Motivates empirical return predictability and analysis

Cross-Asset Relationships

- **Correlation Analysis:**
  - Equity–bond correlation occasionally flips sign (*risk-on/risk-off* regimes).
  - FX movements sometimes reflect broader risk sentiment correlated with equities.
- **Volatility Patterns:**
  - Commodity volatility often spikes in response to geopolitical events or supply shocks.
  - FX volatility partially driven by monetary policy divergence among central banks.
- **Preliminary Takeaway:**
  - Each asset class truly has **distinct** drivers, justifying our multi-agent approach.
  - Aligns with the literature: different short-term vs. long-term predictors in each market.

Data Collection and Features (FX as an example)

- **Currency ETFs**: FXE, FXB, FXY, FXF, FXC
- **Momentum Features**:
  - 1-month, 3-month, 12-month percentage returns
- **Risk Sentiment Indicators**:
  - VIX and MOVE indices with weekly changes
- **Macro Rates**:
  - 10Y government bond yields from top 5 currency pair (US vs. UK, EU, JP, CH, CA)
  - Weekly rate changes calculated
- Data saved at `weekly` frequency for prediction.

LLM Output Schema for Predictions

Structured JSON Schema:

- `currency_pairs` **(array of objects)**:
  - `pair`: one of [EUR/USD, GBP/USD, USD/JPY, USD/CHF, USD/CAD]
  - `predicted_return`: float (e.g., 0.0025 for +0.25%)
  - `confidence`: float in [0, 1]
  - `rationale`: short justification
- `overall_analysis` **(string)**:
  - Summary of market conditions, cross-currency drivers, risk sentiment

Purpose: Enables structured financial forecasting, ensuring interpretability and ease of downstream analysis.

LLM Prompt Template: Weekly Forecast (Forex)

Instruction to GPT-4o:

"Based on this week's market data (ETF returns, 10Y rates, VIX/MOVE), predict next week's return, confidence, and rationale for each FX pair. Also include overall market analysis. Use the required JSON schema."

Market Features Included:

- **Momentum** (1m, 3m, 12m) of 5 Currency ETFs (FXE, FXB, FXY, FXF, FXC)
- **10Y Sovereign Yields** (level and weekly change for US, EUR, GBP, JPY, CHF, CAD)
- **Risk Indicators:** VIX and MOVE index + weekly deltas

Output: Deterministic (temperature = 0), formatted strictly to JSON schema.

LLM Prompt Template: Weekly Forecast (Fixed Income)

Instruction to GPT-4o:

"Based on this week's fixed income market data (macro indicators, Treasury yields, momentum signals, risk sentiment, and Treasury ETF returns), predict next week's yield changes and returns for each Treasury ETF instrument. Include overall market analysis. Use the required JSON schema."

Market Features Included:

- **Macro Indicators:** EFFR, Headline PCE, Core PCE.
- **US Treasury Yields and Momentum:** Yields and momentum (1m, 3m, 12m) for 3-Month, 6-Month, 1-Year, 2-Year, 5-Year, and 10-Year Treasuries.
- **ETF Returns:** Short-Term Treasury (SHV), 1-3 Year Treasury (SHY), 3-7 Year Treasury (IEI), 7-10 Year Treasury (IEF), 10-20 Year Treasury (TLH), and 20+ Year Treasury (TLT).

Output: Deterministic (temperature = 0), formatted strictly to JSON schema

Pipeline Predictions Snapshot

ETF	Instrument	Predicted	Conf.	Rationale
SHV	Short-Term Treasury	0.01	0.8	With the effective federal funds rate remaining high at 5.33%, short-term yields are likely to remain stable, leading to modest returns.
SHY	1-3 Year Treasury	0.009	0.75	The 1-3 year segment is expected to see stable yields due to the current economic environment, with slight upward pressure from the Fed's stance.
IEF	7-10 Year Treasury	0.005	0.65	The 7-10 year segment may experience slight yield increases as investors adjust to inflation expectations, but overall stability is anticipated.

LLM Prediction Performance

Currency Pair	Avg Confidence	Directional Accuracy
USD/CAD	0.6281	0.7260
USD/CHF	0.5897	0.6849
USD/JPY	0.6644	0.6027
EUR/USD	0.7192	0.4795
GBP/USD	0.6658	0.4658
Overall	—	**0.5918**

Note: USD-cross pairs such as USD/CAD, USD/CHF, and USD/JPY outperform others in confidence and directional accuracy.

Confidence vs Accuracy by Currency Pair

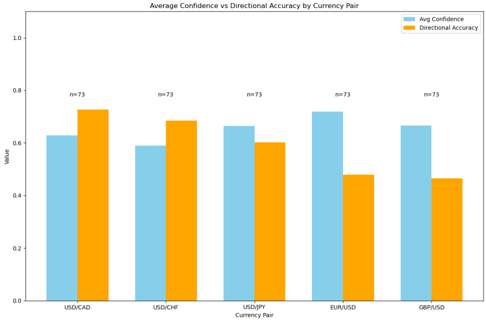


Figure 2: *

Confidence and directional accuracy compared across currency pairs.

Directional Accuracy by Model Confidence

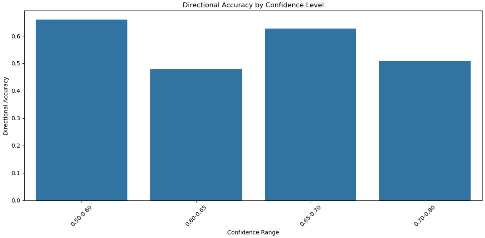


Figure 3: *

Model confidence is not always aligned with predictive accuracy. Surprisingly, predictions in the 0.50–0.60 range are more accurate than higher-confidence bins.

Black-Litterman Portfolios: Fixed Income and Forex

Fixed Income Portfolio (BL vs. Equal-Weighted)

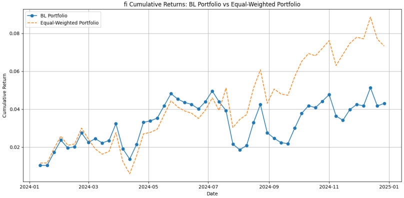


Figure 4: Cumulative Returns: BL Portfolio vs. Equal-Weighted (Fixed Income)

Metric	Value
Annualized Return	3.60 %
Annualized Std Dev	4.72 %
Sharpe Ratio	0.7626
Max Drawdown	-0.0296
Average Turnover	0.3403

Table 2: Performance Metrics: Fixed Income BL Portfolio

Black-Litterman Portfolios: Fixed Income and Forex

Forex Portfolio (BL vs. Equal-Weighted)

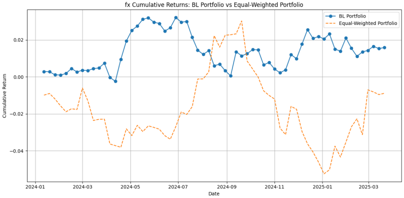


Figure 5: Cumulative Returns: BL Portfolio vs. Equal-Weighted (Forex)

Metric	Value
Annualized Return	1.31 %
Annualized Std Dev	3.32 %
Sharpe Ratio	0.3937
Max Drawdown	-0.0305
Average Turnover	0.7681

Table 3: Performance Metrics: Forex BL Portfolio

Interim Result Analysis

Both agent portfolios exhibit lower but more stable returns.

1. **LLM Forecast Conservatism:**
   - The Large Language Model (LLM) forecasts are consistently cautious and rarely predict large anomalies or outliers.
   - This aligns with the tendency of LLMs to avoid extreme or unverified conclusions (lack of "critical evaluation").
2. **Black-Litterman Constraints:**
   - The quadratic constraints and regularization in the BL framework inherently dampen aggressive weight shifts.
   - This systematic "smoothing" leads to lower volatility in returns.

The combination of conservative LLM-driven forecasts and the weight-stabilizing effect of the Black-Litterman approach may naturally result in portfolios that prioritize stability over high returns.

Next Steps in Our Project Timeline

- **Finalize Agent Pipeline:**
  - Feature Engineering: Apply EWMA and ZScore for more robust return estimation
  - Prompt Engineering: guided two-shot prompting with analytical sample.
  - Incorporate return estimation model to help ensure consistency in agent prediction.
  - Add ToolCalling for agent to get real-time news and chain-of-thought sentiment analysis with SerpAPI
  - Fine-tune model specifications, explore ReAct capabilities for the agent to explore secondary follow-up action for analysis.
- **Document & Validate Model:**
  - Compare ABBL performance to standard momentum strategies and factor-based approaches.
  - Develop risk metrics (volatility, VaR, CVaR) for the final portfolio (potentially risk manager agent).

Areas for Further Research (1/2)

- **Agent Sentiment, ToolCalling & MCP:**
  - Integrate advanced large language models with ToolCalling (e.g., via SerpAPI) to continuously scan and parse high-impact market events such as FOMC testimonies, tariff announcements, and geopolitical developments.
  - Leverage the extracted sentiment signals to update the agent's memory, refine real-time portfolio forecasts, and utilize a greater access to tools and API through Model-Context-Protocol (MCP) for enhanced decision-making.
- **Regime Shifts & Nonlinearities:**
  - Markov Switching or threshold models for detecting structural breaks.
  - Agents can adapt to high-volatility or crisis regimes differently than stable periods.

Areas for Further Research (2/2)

- **Reinforcement Learning Extensions:**
  - Explore dynamic multi-agent manager that evaluate and regulate asset-specifc agents.
  - Model fine-tuning through reinforcement learnings.
- **Robustness Checks & Causal Identification:**
  - Test final model across different market conditions (bull vs. bear).
  - Investigate causal drivers vs. correlation-based signals.
  - Stress testing and sensitivity analysis to validate risk management.

Additional Possible Directions

Potential directions for the next step could include:

1. **Use of a standard model as the main "absolute view":** Incorporate agents as providers of *relative views* (i.e., identifying which assets might outperform others) based on data and real-time news feeds.
2. **Fine-tune a base model with reinforcement learning:** Aim to train an advanced LLM or "tradingAgent" setup; let agents debate and exchange signals in a dynamic environment to improve forecasting accuracy.
3. **User-Focused Extensions:** Similar to FolioLLM, integrate natural language input to gather user constraints and preferences. Generate straightforward reports and then apply a standard Markowitz procedure (with risk aversion parameters derived from user inputs).
4. **Expanded Data Sourcing:** Explore more comprehensive data offerings from Bloomberg or other vendors to enhance coverage an d accuracy of each asset's forecast.

Conclusion

- **Key Takeaways:**
  - We combine **multi-asset, agent-based forecasting** with a **Bayesian Black-Litterman** framework.
  - The approach is flexible, scalable, and focuses on capturing heterogeneous market signals.
  - We believe this can lead to better risk-adjusted performance and adaptability to market regime changes.
- **Looking Ahead:**
  - Incorporate final empirical evidence and robust testing.
  - Potentially expand into LLM-based macro and sentiment forecasting.

Reference

Ang, A. (2014). *Asset management*. Oxford University Press.

Black, F., & Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal*, *48*(5), 28–43.

Farmer, J. D., & Foley, D. (2009). The economy needs agent-based modelling. *Nature*, *460*(7256), 685–686.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56.

He, G., & Litterman, R. (1999). *The intuition behind the Black-Litterman model portfolios*. Goldman Sachs Asset Management.

Reference-2

LeBaron, B. (2006). Agent-based computational finance. In *Handbook of Computational Economics* (Vol. 2, pp. 1187–1233). North-Holland.

Li, X., Gao, Y., & Li, Z. (2024). EconAgent: LLM-powered agent-based macroeconomic modeling. *Working paper*.

Popov, A., & Roshka, D. (2024). FolioLLM: A domain-specific large language model for ETF portfolio construction. *Working paper*.

Satchell, S., & Scowcroft, A. (2000). A demystification of the Black-Litterman model: Managing quantitative and traditional portfolio construction. *Journal of Asset Management*, *1*(2), 138–150.

Wu, S., Li, J., & Tan, P. (2023). AI-driven sentiment analysis for market prediction. *Journal of Financial Technology*, *12*(3), 45–63.