Capstone Project: Complete Bayesian Forecas System for Commodities

Course: Bayesian Regression and Time Series Forecasting for Commodities Trading

---

Project Overview

In this capstone project, you will build a **complete end-to-end Bayesian forecasting and tradi system** for a portfolio of commodities. This project integrates all concepts learned throughout th

- Bayesian inference and prior selection
- Time series analysis and decomposition
- Bayesian regression and structural time series
- Hierarchical models for multiple assets
- Volatility modeling and risk management
- Proper backtesting and evaluation
- Trading strategy implementation

Project Requirements

Deliverables

1. **Data Pipeline**: Download and preprocess at least 3 - 5 related commodities
2. **Exploratory Analysis**: Time series characteristics, seasonality, correlations
3. **Bayesian Models**: At least 2 different model types with proper diagnostics
4. **Forecasting System**: Multi-step probabilistic forecasts
5. **Trading Strategy**: Uncertainty-aware position sizing
6. **Backtesting**: Walk-forward validation with proper metrics
7. **Risk Analysis**: VaR, CVaR, drawdown analysis
8. **Written Report**: 5 -page summary with key findings

Grading Rubric

Component	Weight	Criteria
Technical Implementation	40%	Code quality, model sophistication, diagnostics
Performance	30%	Forecast accuracy (CRPS), trading metrics (Sharpe)
Risk Management	20%	Uncertainty quantification, position sizing
Documentation	10%	Clear explanation, reproducibility

Part 1 : Data Pipeline and Exploratory Analysis

1.1 Select Your Commodity Portfolio

Choose ONE of these commodity complexes:

Option A: Energy Complex

- WTI Crude Oil (CL=F)
- Brent Crude Oil (BZ=F)
- Natural Gas (NG=F)
- Gasoline (RB=F)
- Heating Oil (HO=F)

Option B: Agricultural Complex

- Corn (ZC=F)
- Wheat (ZW=F)
- Soybeans (ZS=F)
- Soybean Oil (ZL=F)
- Soybean Meal (ZM=F)

Option C: Precious Metals

- Gold (GC=F)
- Silver (SI=F)
- Platinum (PL=F)
- Copper (HG=F)

Option D: Custom (subject to approval)

- Choose 3 - 5 related commodities with economic rationale

In [ ]:
```
# Setup
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from scipy import stats
import pymc as pm
import arviz as az
import warnings
warnings.filterwarnings('ignore')

# Course utilities
import sys
sys.path.append('..')
from datasets.download_data import (
    get_commodity_data,
    get_multiple_commodities,
    prepare_commodity_dataset,
    create_train_test_split
)
from utils.plotting import (
```

```
    plot_time_series,
    plot_forecast,
    plot_posterior,
    plot_backtest_results
)
from utils.metrics import (
    crps, crps_gaussian,
    forecast_summary,
    backtest_summary
)
from utils.backtesting import (
    WalkForwardValidator,
    Backtester,
    create_folds
)

np.random.seed(42)
plt.style.use('seaborn-v0_8-whitegrid')

print("Setup complete!")
```

In [ ]:
```
# TODO: Define your commodity portfolio
# Example for Energy Complex:

COMMODITIES = [
    'crude_oil',
    'natural_gas',
    'gasoline',
    # Add more commodities...
]

START_DATE = '2015-01-01'
END_DATE = None  # Today

# Download data
# YOUR CODE HERE
data = get_multiple_commodities(COMMODITIES, start=START_DATE, end=END_DA
print(f"Downloaded {len(data.columns)} commodities")
print(f"Date range: {data.index[0]} to {data.index[-1]}")
print(f"Total observations: {len(data)}")
data.head()
```

In [ ]:
```
# TODO: Exploratory Data Analysis
# 1. Plot all commodity prices (normalized)
# 2. Calculate and visualize correlations
# 3. Test for stationarity (ADF, KPSS)
# 4. Analyze seasonality patterns
# 5. Identify any structural breaks or regime changes

# YOUR CODE HERE
```

1.2 Document Your Findings

Questions to Answer:

1. What are the key characteristics of your chosen commodities?
2. Are there clear seasonal patterns? Which commodities show strongest seasonality?
3. What is the correlation structure? Are there natural pairs for spread trading?

4. Are the series stationary? What transformations are needed?
5. Are there any obvious regime changes in the data?

Your EDA Summary Here:

-
-
-

Part 2: Bayesian Forecasting Models

Implement at least TWO of the following model types:

1. **Bayesian Linear Regression** with economic factors
2. **Bayesian Structural Time Series** (BSTS)
3. **Hierarchical Model** for multiple commodities
4. **Gaussian Process** regression
5. **Stochastic Volatility** model

For each model:

- Justify your prior choices
- Run prior predictive checks
- Fit the model with PyMC
- Check convergence (R-hat, ESS, trace plots)
- Run posterior predictive checks
- Generate forecasts with uncertainty

2.1 Model 1: [Your Choice]

Model Description: [Explain your model choice and why it's appropriate]

In [ ]:
```
# Model 1 Implementation
#
# TODO:
# 1. Define priors with justification
# 2. Build PyMC model
# 3. Run prior predictive check
# 4. Fit model (sample from posterior)
# 5. Check convergence
# 6. Run posterior predictive check

# YOUR CODE HERE
```

2.2 Model 2: [Your Choice]

Model Description: [Explain your model choice and why it's appropriate]

In [ ]:
```

```

```
# Model 2 Implementation
#
# YOUR CODE HERE
```

2.3 Model Comparison

Compare your models using:

- WAIC (Widely Applicable Information Criterion)
- LOO (Leave-One-Out Cross-Validation)
- Out-of-sample CRPS

In [ ]:
```
# Model Comparison
#
# YOUR CODE HERE
```

Part 3: Trading Strategy Development

Implement a trading strategy that:

1. Uses your Bayesian forecasts
2. Accounts for forecast uncertainty in position sizing
3. Manages risk appropriately

Choose from:

- **Mean reversion** with credible interval bands
- **Trend following** with probability-based entries
- **Pairs/spread trading** with hierarchical model
- **Volatility targeting** with stochastic vol forecasts
- **Custom strategy** (describe your approach)

In [ ]:
```
# Trading Strategy Implementation
#
# TODO:
# 1. Define signal generation from forecasts
# 2. Implement position sizing (Kelly criterion or similar)
# 3. Add risk management rules
# 4. Generate signals for full history

def generate_trading_signals(forecasts, actuals, model_uncertainty):
    """
    Generate trading signals from Bayesian forecasts.

    Parameters:
    -----------
    forecasts : pd.DataFrame
        Posterior predictive samples or point forecasts
    actuals : pd.Series
        Actual price series
    model_uncertainty : pd.Series
        Posterior standard deviation
```

```
    Returns:
    --------
    pd.Series
        Signal series (-1 to 1)
    """
    # YOUR CODE HERE
    pass

def bayesian_position_size(signal, forecast_mean, forecast_std,
                           max_position=1.0, kelly_fraction=0.5):
    """
    Calculate position size using Bayesian Kelly criterion.

    Parameters:
    -----------
    signal : float
        Trading signal direction
    forecast_mean : float
        Expected return
    forecast_std : float
        Standard deviation of forecast
    max_position : float
        Maximum position size
    kelly_fraction : float
        Fraction of Kelly (0.5 = half-Kelly)

    Returns:
    --------
    float
        Position size
    """
    # YOUR CODE HERE
    pass
```

Part 4 : Backtesting and Evaluation

Perform rigorous backtesting with:

1. Walk-forward validation (no look-ahead bias)
2. Proper evaluation metrics (CRPS for forecasts, Sharpe for trading)
3. Comparison to benchmark strategies

In [ ]:
```
# Walk-Forward Backtesting
#
# TODO:
# 1. Set up walk-forward validation folds
# 2. For each fold: fit model, generate forecasts, calculate signals
# 3. Compute evaluation metrics per fold
# 4. Aggregate results

# YOUR CODE HERE
```

In [ ]:
```

```

```
# Full Backtest
#
# YOUR CODE HERE
```

In [ ]:
```
# Performance Summary
#
# TODO: Create comprehensive performance report

# YOUR CODE HERE
```

Part 5 : Risk Analysis

Analyze the risk characteristics of your strategy:

1. Value at Risk (VaR) and Conditional VaR (CVaR)
2. Maximum drawdown analysis
3. Tail risk assessment
4. Scenario analysis (what if volatility doubles?)

In [ ]:
```
# Risk Analysis
#
# YOUR CODE HERE
```

Part 6 : Final Report

Write a 5 -page summary covering:

1. **Executive Summary** (0.5 page)
    - Key findings and performance metrics
2. **Data and Methodology** (1 page)
    - Commodity selection rationale
    - Model choices and prior justification
3. **Model Results** (1.5 pages)
    - Posterior analysis
    - Forecast accuracy
    - Model comparison
4. **Trading Performance** (1 page)
    - Backtest results
    - Risk metrics
    - Comparison to benchmarks
5. **Conclusions and Limitations** (1 page)
    - Key insights

- What worked, what didn't
- Suggestions for improvement

Your Report Here

(Use markdown formatting for your written report)

---

1. Executive Summary

[Your summary here]

---

2. Data and Methodology

[Your methodology here]

---

3. Model Results

[Your results here]

---

4. Trading Performance

[Your performance analysis here]

---

5. Conclusions and Limitations

[Your conclusions here]

Submission Checklist

Before submitting, ensure you have:

- ☐ Downloaded and preprocessed at least 3 commodities
- ☐ Completed exploratory data analysis with visualizations
- ☐ Implemented at least 2 Bayesian models with proper diagnostics
- ☐ Justified all prior choices
- ☐ Generated probabilistic forecasts
- ☐

Implemented a trading strategy with uncertainty-aware position sizing

- ☐ Performed walk-forward backtesting
- ☐ Calculated proper evaluation metrics (CRPS, Sharpe, etc.)
- ☐ Completed risk analysis (VaR, drawdown, etc.)
- ☐ Written 5 -page summary report
- ☐ All code runs without errors
- ☐ Results are reproducible (random seed set)

Good luck! This project is your opportunity to demonstrate mastery of Bayesian forecasti commodity trading.