Kalman Filter and Its Application in Pairs Trading

Nimay Shah and Devavrat Patni

Contents

1 Introduction **2**

2 The Kalman Filter **2**
2.1 Mathematical Formulation 2
2.1.1 Prediction Step 2
2.1.2 Update Step 3
2.2 Derivation and Intuition 3

3 Pairs Trading Using Kalman Filter **4**
3.1 Key Steps in Pairs Trading 4
3.2 Hurst Test 4
3.3 ADF Test 6
3.4 Half-Life of Mean Reversion 6
3.5 Kalman Filter for Pairs Trading 6
3.5.1 Kalman Filter Average 6
3.5.2 Kalman Filter Regression 6

4 Code Implementation and Explanation **9**
4.1 KalmanFilterAverage 9
4.2 KalmanFilterRegression 9
4.3 Backtest Function 9
4.4 Sharpe Ratio 10
4.5 Backtesting: In-Sample and Out-of-Sample Testing 11
4.6 Drawdown in Backtesting 12

5 Observations **13**

6 Conclusion **14**

7 Next Up **14**

1 Introduction

The Kalman Filter is a recursive algorithm used for estimating the state of a dynamic system from noisy observations. It has wide applications in fields such as signal processing, control systems, and finance. In the context of pairs trading, the Kalman Filter is used to model and estimate the dynamic relationship between two financial assets, providing a robust framework for identifying trading signals.

2 The Kalman Filter

2.1 Mathematical Formulation

The Kalman Filter operates on a linear dynamic system described by the following equations:

$$x_k = Ax_{k-1} + w_{k-1}, \quad w_{k-1} \sim \mathcal{N}(0, Q), \tag{1}$$
$$z_k = Hx_k + v_k, \quad v_k \sim \mathcal{N}(0, R), \tag{2}$$

where:

- x_k is the state vector at time k.
- z_k is the observation vector at time k.
- A is the state transition matrix.
- H is the observation matrix.
- w_{k-1} is the process noise with covariance Q.
- v_k is the measurement noise with covariance R.

These equations originate from modeling the evolution of a system (state transition equation) and the relationship between the system's state and observed measurements (observation equation). Process noise (w_{k-1}) represents uncertainty in system dynamics, while measurement noise (v_k) accounts for observational inaccuracies.

2.1.1 Prediction Step

The state and its covariance are predicted as:

$$\hat{x}_{k|k-1} = A\hat{x}_{k-1|k-1}, \tag{3}$$
$$P_{k|k-1} = AP_{k-1|k-1}A^T + Q, \tag{4}$$

where $\hat{x}_{k|k-1}$ is the predicted state, and $P_{k|k-1}$ is the predicted covariance. These equations propagate the state forward in time and account for uncertainties introduced by process noise.

2.1.2 Update Step

The state and covariance are updated using the observation z_k:

$$K_k = P_{k|k-1} H^T (H P_{k|k-1} H^T + R)^{-1}, \tag{5}$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k (z_k - H \hat{x}_{k|k-1}), \tag{6}$$

$$P_{k|k} = (I - K_k H) P_{k|k-1}, \tag{7}$$

where K_k is the Kalman Gain. The Kalman Gain determines the weighting between the prediction and the new observation, ensuring that more accurate measurements contribute more to the state estimate.

2.2 Derivation and Intuition

The Kalman Filter equations can be derived from Bayesian principles. The goal is to estimate the posterior probability distribution of the state x_k given the observations up to time k, $z_{1:k}$. The prediction step estimates $p(x_k|z_{1:k-1})$, while the update step computes $p(x_k|z_{1:k})$ using Bayes' theorem.

Intuitively:

- The prediction step propagates the state forward, assuming the system dynamics described by A.
- The update step corrects this prediction using the new observation, balancing the uncertainties in the prediction ($P_{k|k-1}$) and measurement noise (R).

3 Pairs Trading Using Kalman Filter

Pairs trading involves identifying two assets with a stable statistical relationship, such as cointegration, and exploiting deviations from this relationship. The Kalman Filter is used to estimate the time-varying parameters of this relationship in real time.

3.1 Key Steps in Pairs Trading

3.2 Hurst Test

The Hurst test is a statistical method used to evaluate the degree of mean reversion or persistence in a time series. It is named after Harold Edwin Hurst, who originally used it to analyze long-term storage capacities of reservoirs. In financial applications, the Hurst exponent provides insight into the behavior of asset prices, which can be leveraged to design trading strategies, including pairs trading.

Mathematical Formulation

The Hurst exponent H is derived from the relationship between the time lag l and the rescaled range of the time series. It is computed as:

$$\tau(l) = \sqrt{\mathrm{Var}(X_{t+l} - X_t)}, \tag{8}$$

where $\tau(l)$ is the expected value of the absolute differences for a lag l. The Hurst exponent is estimated as the slope of the line in a log-log plot of $\tau(l)$ versus l:

$$H = \frac{\log(\tau(l))}{\log(l)}. \tag{9}$$

Interpretation of Hurst Exponent

The value of the Hurst exponent H characterizes the behavior of the time series.
Intuition: When $H < 0.5$, the time series exhibits negative autocorrelation. This means that an increase (or decrease) in the value of the series is more likely to be followed by a decrease (or increase), causing the series to revert toward its mean. This behavior is a key property for identifying pairs in pairs trading, as it indicates that the spread between two assets is likely to oscillate around a stable value.

- $H < 0.5$: The time series exhibits mean-reverting behavior, which is a crucial property for pairs trading.
- $H = 0.5$: The time series resembles a random walk (no mean reversion or trend persistence).
- $H > 0.5$: The time series exhibits persistence or trending behavior.

Relevance to Pairs Trading

In the context of pairs trading, the Hurst exponent is applied to the spread between two assets. A low Hurst exponent ($H < 0.5$) indicates that the spread is mean-reverting, suggesting that deviations from the mean are likely to reverse, making it a candidate for pairs trading. This is particularly useful for pairs trading because:

- **Mean-Reverting Nature:** A low H implies that deviations from the mean are not persistent and are likely to reverse. This provides opportunities to buy when the spread is below the mean and sell when it is above the mean.
- **Statistical Arbitrage:** The reversion behavior ensures that the spread is not random but follows a predictable pattern, enabling traders to capitalize on deviations.
- **Stable Relationship:** Mean reversion suggests a stable statistical relationship between the two assets, making them a reliable candidate for pairs trading.

In summary, a low Hurst exponent indicates that the spread is not dominated by noise or trends but exhibits systematic reversals around a mean, which is a desirable characteristic for pairs trading strategies. For example:

$$\text{Spread: } s_t = y_t - \beta x_t, \tag{10}$$

where y_t and x_t are the price series of two assets and β is the hedge ratio.

What is the Hedge Ratio?

The hedge ratio is a measure used in pairs trading to determine the optimal proportion of one asset to trade against another to minimize risk. Mathematically, it is defined as:

$$\text{Hedge Ratio} = \beta_t \tag{11}$$

where β_t is the time-varying coefficient that captures the sensitivity of the dependent asset (y_t) to changes in the independent asset (x_t).

Intuition:

- A hedge ratio of β_t implies that for every unit of asset x_t, you should hold β_t units of asset y_t to hedge against fluctuations in their spread.
- This ratio ensures that the portfolio is neutral with respect to market movements, allowing traders to profit from deviations in the spread rather than directional market trends.

Steps for Estimating the Hurst Exponent

1. Compute the differences $X_{t+l} - X_t$ for various time lags l.
2. Calculate $\tau(l) = \sqrt{\text{Var}(X_{t+l} - X_t)}$ for each lag.
3. Perform a linear regression on the log-log plot of $\tau(l)$ versus l to estimate the slope, which is the Hurst exponent.

Application Example

Consider a time series s_t representing the spread of two cointegrated assets. By estimating the Hurst exponent:

- If $H < 0.5$, the spread is likely mean-reverting, providing opportunities for profitable trades when the spread deviates from its mean.
- If $H \geq 0.5$, the spread is less suitable for pairs trading as it may exhibit random or trending behavior.

Comparison with Other Stationarity Tests

While the Hurst test provides information on mean reversion, it complements other tests like the Augmented Dickey-Fuller (ADF) test. The ADF test specifically evaluates stationarity, whereas the Hurst test focuses on the persistence or reversion properties of the series.

Limitations

The Hurst test assumes consistent behavior across time scales, which may not always hold for financial time series. Additionally, the presence of noise and short-term trends can affect its accuracy.

3.3 ADF Test

The ADF test was covered in the last document, refer that for the same.

3.4 Half-Life of Mean Reversion

The half-life of mean reversion quantifies how quickly the spread reverts to its mean. It is computed as:

$$\text{Half-life} = \frac{-\ln(2)}{\hat{\phi}}, \tag{12}$$

where $\hat{\phi}$ is the coefficient of s_{t-1} in the regression:

$$\Delta s_t = \alpha + \phi s_{t-1} + \epsilon_t. \tag{13}$$

3.5 Kalman Filter for Pairs Trading

3.5.1 Kalman Filter Average

The Kalman Filter can be used to compute a rolling mean of the price series x_t:

$$\hat{x}_t = \frac{P_t H_t^T}{H_t P_t H_t^T + R} z_t. \tag{14}$$

3.5.2 Kalman Filter Regression

To model the dynamic relationship between two assets x_t and y_t, the Kalman Filter estimates the parameters α_t and β_t in:

$$y_t = \alpha_t + \beta_t x_t + \epsilon_t. \tag{15}$$

The Kalman Filter tracks the parameters α_t and β_t through a state vector:

$$\theta_t = \begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix}, \tag{16}$$

which evolves over time based on observed data.

The filter equations are:

$$\hat{\theta}_{t|t-1} = A\hat{\theta}_{t-1|t-1}, \tag{17}$$

$$\hat{\theta}_{t|t} = \hat{\theta}_{t|t-1} + K_t(y_t - H_t\hat{\theta}_{t|t-1}). \tag{18}$$

where:

- $\hat{\theta}_{t|t-1}$: Predicted state vector $[\alpha_t, \beta_t]^T$.
- $P_{t|t-1}$: Predicted covariance matrix of θ_t.
- A: State transition matrix, defining the dynamics of θ_t (e.g., random walk or stationary process).
- Q: Process noise covariance matrix, representing uncertainty in parameter evolution.

The filter refines its predictions using the latest observations x_t and y_t:

$$K_t = P_{t|t-1}H_t^T\left(H_tP_{t|t-1}H_t^T + R\right)^{-1}, \tag{19}$$

$$\hat{\theta}_{t|t} = \hat{\theta}_{t|t-1} + K_t\left(y_t - H_t\hat{\theta}_{t|t-1}\right), \tag{20}$$

$$P_{t|t} = (I - K_tH_t)\,P_{t|t-1}, \tag{21}$$

where:

- K_t: Kalman gain, weighting the correction from the new observation.
- $H_t = \begin{bmatrix}1 & x_t\end{bmatrix}$: Observation matrix, linking the state vector θ_t to the observation y_t.
- R: Observation noise covariance, representing the variance of the residual ϵ_t.
- $\hat{\theta}_{t|t}$: Updated state vector $[\alpha_t, \beta_t]^T$.
- $P_{t|t}$: Updated covariance matrix.

The parameters are recursively updated as follows:

$$\hat{\theta}_{t|t-1} = A\hat{\theta}_{t-1|t-1}, \tag{22}$$

$$\hat{\theta}_{t|t} = \hat{\theta}_{t|t-1} + K_t\left(y_t - H_t\hat{\theta}_{t|t-1}\right), \tag{23}$$

where the first equation predicts the next state based on a specified dynamic model, and the second equation refines these predictions using new observations. This recursive framework ensures that the estimates of α_t and β_t remain responsive to changing market conditions.

Why is the Hedge Ratio Equal to β_t?

The hedge ratio in pairs trading is determined by the dynamic linear regression model:

$$y_t = \alpha_t + \beta_t x_t + \epsilon_t \tag{24}$$

Here, β_t represents the slope of the regression line, which indicates the relative movement of y_t with respect to x_t.

Reasoning:

- β_t is the time-varying sensitivity coefficient, capturing how changes in x_t affect y_t.
- In pairs trading, the goal is to neutralize the spread $y_t - \beta_t x_t - \alpha_t$, and the hedge ratio β_t minimizes the variance of this spread.
- As β_t dynamically adjusts to changes in the relationship between x_t and y_t, it acts as the optimal hedge ratio at any given time.

The Kalman Filter is a recursive algorithm that estimates the dynamic relationship between two assets in real time. It helps in pairs trading by:

1. **Dynamic Estimation of Parameters:** The filter estimates time-varying coefficients α_t and β_t in the regression model:

   $$y_t = \alpha_t + \beta_t x_t + \epsilon_t \tag{25}$$

   These parameters adapt to changes in the market, ensuring that the hedge ratio remains optimal.

2. **Spread Calculation:** Using the estimated parameters, the spread is calculated as:

   $$\text{Spread}_t = y_t - \beta_t x_t - \alpha_t \tag{26}$$

   The Kalman Filter ensures that this spread reflects the latest market dynamics.

3. **Mean Reversion Detection:** The Kalman Filter detects mean-reverting behavior in the spread by analyzing its variance and persistence, enabling traders to identify profitable entry and exit points.

4 Code Implementation and Explanation

4.1 KalmanFilterAverage

The `KalmanFilterAverage` function uses a Kalman filter to compute a smoothed rolling mean for a given time series x.

- **Transition Matrices**: Controls how the state (mean) evolves over time. Here, it assumes a constant mean with [1].
- **Observation Matrices**: Relates the observed data to the state. This is set to [1], indicating a direct observation of the mean.
- **Initial State Mean**: Assumes the initial state starts at 0.
- **Transition and Observation Covariance**: Define the noise in the process and observations, influencing how much the smoothed mean adapts to changes.

The function outputs a series of rolling mean estimates.

4.2 KalmanFilterRegression

The `KalmanFilterRegression` function applies a Kalman filter to dynamically estimate the relationship between two time series, x and y, in the form:

$$y_t = \alpha_t + \beta_t x_t + \epsilon_t,$$

where α_t and β_t evolve over time.

- **State Parameters**: Includes α_t (intercept) and β_t (slope or hedge ratio), making the state dimension 2.
- **Transition Matrices**: Assumes the state parameters evolve independently with random noise.
- **Observation Matrices**: Models y_t as a function of x_t and a constant term.
- **Transition and Observation Covariance**: Control the noise in the state evolution and observations, allowing flexibility in adapting to data changes.

The function returns time-varying estimates of α_t and β_t.
—

4.3 Backtest Function

The `backtest` function evaluates the performance of a pairs trading strategy using the Kalman filter for hedge ratio estimation.

- **Inputs**:
  - $s1, s2$: Symbols of the two assets.
  - x, y: Price series of the two assets.

- **Key Steps**:
  1. **Hedge Ratio Calculation**: Uses the `KalmanFilterRegression` to dynamically compute the hedge ratio (β_t).
  2. **Spread Calculation**: The spread is defined as:
  $$\text{spread}_t = y_t + \beta_t x_t.$$
  3. **Z-Score Computation**: Normalizes the spread using its rolling mean and standard deviation:
  $$z_t = \frac{\text{spread}_t - \mu_t}{\sigma_t}.$$
  4. **Trading Rules**: Signals are generated based on the Z-score:
     - **Long Entry**: When $z_t < -2$.
     - **Short Entry**: When $z_t > 2$.
     - **Exit**: When z_t reverts to 0.

     this is similar to what was done previously, check out the last document if you don't follow this.
  5. **Returns and Sharpe Ratio**: Calculates cumulative portfolio returns and the Sharpe ratio for performance evaluation:
  $$\text{Sharpe} = \frac{\mu_{\text{returns}}}{\sigma_{\text{returns}}}\sqrt{252},$$
  where 252 is the number of trading days in a year.

Outputs:

- **Cumulative Returns**: The compounded returns from the strategy.
- **Sharpe Ratio**: A measure of risk-adjusted returns.

4.4 Sharpe Ratio

The Sharpe Ratio is a widely used metric in finance to measure the performance of an investment or trading strategy relative to its risk. It is defined as:

$$\text{Sharpe Ratio} = \frac{R_p - R_f}{\sigma_p},$$

where:

- R_p: The mean return of the portfolio or strategy.
- R_f: The risk-free rate of return, typically the return of government bonds.
- σ_p: The standard deviation of the portfolio's returns, representing the risk or volatility.

Interpretation

- A higher Sharpe Ratio indicates better risk-adjusted returns, meaning the strategy generates higher excess returns per unit of risk.
- A negative Sharpe Ratio implies that the strategy underperforms the risk-free rate.

Annualized Sharpe Ratio

In the backtest function, the Sharpe Ratio is annualized using:

$$\text{Annualized Sharpe Ratio} = \frac{\mu_{\text{returns}}}{\sigma_{\text{returns}}}\sqrt{T},$$

where:

- μ_{returns}: The mean daily returns of the strategy.
- σ_{returns}: The standard deviation of the daily returns.
- T: The number of trading days in a year (typically $T = 252$).

Limitations

While the Sharpe Ratio is a useful metric, it has certain limitations:

- It assumes returns are normally distributed, which may not always be the case.
- It does not account for the impact of extreme events or non-linear risks.

Despite these limitations, the Sharpe Ratio remains a standard tool for evaluating risk-adjusted performance in finance.

4.5 Backtesting: In-Sample and Out-of-Sample Testing

Backtesting is a critical process in financial modeling and algorithmic trading used to evaluate the performance of a strategy using historical data. It allows practitioners to understand how a model or strategy would have performed in the past, providing insights into its potential future performance. However, it is essential to follow robust practices, such as splitting data into *in-sample* and *out-of-sample* sets, to avoid overfitting and ensure realistic results.

In-Sample Testing

In-sample testing involves training and calibrating the model using a subset of historical data. This process helps the model learn patterns, optimize parameters, and fit the data effectively.

- **Purpose:** To develop and fine-tune the strategy.
- **Caution:** Overfitting may occur if the model performs exceptionally well on this data but fails to generalize to unseen data.

Example: A portfolio optimization algorithm is trained on stock prices from 2010 to 2018 to determine optimal weights for assets.

Out-of-Sample Testing

Out-of-sample testing evaluates the model's performance on unseen data that was not used during training. This step is crucial for testing the model's robustness and generalizability.

- **Purpose:** To validate the strategy's effectiveness on new data and avoid overfitting.
- **Caution:** A significant drop in performance from in-sample to out-of-sample indicates overfitting or poor model generalization.

Example: After training on data from 2010 to 2018, the same portfolio optimization algorithm is tested on stock prices from 2019 to 2022.

4.6 Drawdown in Backtesting

In the context of backtesting a pairs trading strategy, drawdown is defined as the largest observed loss from a peak to a trough in the value of a portfolio or trading strategy. Mathematically, the drawdown at time t is given by:

$$\text{Drawdown}(t) = \frac{V_{\text{peak}} - V_t}{V_{\text{peak}}}$$

where:

- V_{peak} is the maximum value of the portfolio up to time t,
- V_t is the value of the portfolio at time t.

The maximum drawdown (MDD) is the largest value of drawdown observed during the backtesting period:

$$\text{MDD} = \max_t \left(\text{Drawdown}(t)\right)$$

In pairs trading, particularly when using Kalman filtering to model the spread between two stocks, the drawdown can be tracked as follows. Let $S_A(t)$ and $S_B(t)$ denote the price of stock A and stock B at time t. The spread $\Delta(t)$ between the two stocks is modeled as:

$$\Delta(t) = S_A(t) - \hat{\beta}(t)S_B(t)$$

where $\hat{\beta}(t)$ is the time-varying coefficient estimated by the Kalman filter. The strategy assumes that the spread $\Delta(t)$ is mean-reverting, and the positions in the stocks are adjusted based on this model.

The value of the portfolio is computed based on the positions in stock A and stock B, and the drawdown can be computed by comparing the portfolio value to its historical peak. In the case of a Kalman filter-based pairs trading strategy, drawdown can reflect the times when the strategy deviates significantly from expected mean-reverting behavior, indicating a potential period of underperformance.

5 Observations

Based on the backtesting results for various commodity pairs, we can draw the following observations:

- **Total Return:** Most pairs showed positive total returns, with *PB-J* achieving the highest return of 26.03%. On the other hand, pairs like *RB-PB* and *AG-RM* performed relatively lower with returns around 10.55% and 11.03%, respectively.
- **Daily Sharpe Ratio:** The pair *RB-TA* has the highest daily Sharpe ratio of 3.02, indicating excellent risk-adjusted returns. Pairs such as *JM-OI* and *RB-PB* have Sharpe ratios below 1, suggesting higher risks relative to their returns.
- **Max Drawdown:** The maximum drawdown values indicate the largest observed loss from a peak to a trough. *AG-M* experienced the smallest drawdown at -1.85%, while pairs like *JM-PB* had significantly higher drawdowns around -10.87%.
- **CAGR (Compound Annual Growth Rate):** The annualized growth rate shows consistent performance, with *PB-J* achieving the highest CAGR of 10.98%, while *RB-PB* showed the lowest at 4.62%.
- **Calmar Ratio:** A high Calmar ratio is desirable, and *RB-TA* stands out with a ratio of 8.84, indicating superior performance relative to drawdown risk. Pairs like *RB-PB* and *JM-PB* have lower ratios, showing less favorable risk-return profiles.
- **Sortino Ratio:** Several pairs, such as *RB-TA* and *PP-NI*, have high Sortino ratios, demonstrating strong performance against downside volatility. However, pairs like *RB-PB* and *JM-PB* need further optimization.

These observations provide insights into the strengths and weaknesses of the strategies for different commodity pairs, enabling informed decisions for future adjustments and optimizations.

6 Conclusion

The Kalman Filter is a powerful tool for real-time estimation in pairs trading. It provides robust parameter estimates, enabling traders to identify and act on profitable opportunities. By integrating statistical tests like the Hurst exponent and ADF test, as well as computing the half-life of mean reversion, the Kalman Filter-based approach ensures a disciplined and systematic trading strategy.

7 Next Up

We make our approach to Kalman Filtering more refined, more structured and more focussed, now attacking the age-old challenge of achieving more returns as well with pair