Stochastic Foundations of Dynamic Pairs Trading

A Bayesian State-Space Approach

Quantitative Research Division

December 21, 2025

Contents

1 Measure-Theoretic Foundations

We operate on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. Let $W = \{W_t\}_{t\geq 0}$ be a standard Brownian motion.

1.1 Quadratic Variation

The cornerstone of stochastic calculus is the non-vanishing quadratic variation of Brownian paths, which necessitates Itô's Lemma.

Theorem 1.1: Quadratic Variation Convergence

Let $\Pi_n = \{t_0, \ldots, t_n\}$ be a partition of $[0, t]$ with mesh $\|\Pi_n\| \to 0$. The quadratic variation converges in $L^2(\Omega)$ (mean-square):

$$\lim_{n\to\infty} \sum_{i=0}^{n-1} (W_{t_{i+1}} - W_{t_i})^2 = t \tag{1}$$

Proof of L^2 Convergence

Let $\Delta W_i \sim \mathcal{N}(0, \Delta t_i)$. We compute the variance of the sum $Q_n = \sum \Delta W_i^2$. Since $\text{Var}(\chi_1^2) = 2$, we have $\text{Var}(\Delta W_i^2) = 2(\Delta t_i)^2$.

$$\text{Var}(Q_n) = \sum_i 2(\Delta t_i)^2 \leq 2 \max(\Delta t_i) \sum \Delta t_i = 2\|\Pi_n\|t$$

As $\|\Pi_n\| \to 0$, $\text{Var}(Q_n) \to 0$. By Chebyshev's inequality, convergence in probability follows. This justifies the differential notation $(dW_t)^2 = dt$.

2 The Ornstein-Uhlenbeck Process

We model the spread ϵ_t via the Langevin Stochastic Differential Equation (SDE):

$$d\epsilon_t = -\theta(\epsilon_t - \mu)dt + \sigma dW_t \tag{2}$$

2.1 Exact Solution and Moments

Lemma 2.1: Explicit Solution via Itô Calculus

The unique strong solution to Eq. (2) is:

$$\epsilon_t = \epsilon_0 e^{-\theta t} + \mu(1 - e^{-\theta t}) + \int_0^t \sigma e^{-\theta(t-u)} dW_u \tag{3}$$

Derivation

Let $f(t, \epsilon) = \epsilon e^{\theta t}$. By Itô's Lemma:

$$df = \left(\frac{\partial f}{\partial t} - \theta(\epsilon - \mu)\frac{\partial f}{\partial \epsilon} + \frac{1}{2}\sigma^2 \frac{\partial^2 f}{\partial \epsilon^2} \right) dt + \sigma \frac{\partial f}{\partial \epsilon} dW_t$$

$$df = \theta \mu e^{\theta t} dt + \sigma e^{\theta t} dW_t$$

Integrating from 0 to t and multiplying by $e^{-\theta t}$ yields the result.

2.2 Stationarity Proof

To prove the strategy is viable, we must prove the variance is bounded (Stationarity).

Theorem 2.1: Conditional Moments & Stationarity

The conditional variance of the process is:

$$\mathrm{Var}(\epsilon_t|\epsilon_0) = \frac{\sigma^2}{2\theta}(1 - e^{-2\theta t}) \tag{4}$$

As $t \to \infty$, $\mathrm{Var}(\epsilon_t) \to \frac{\sigma^2}{2\theta} < \infty$.

Proof via Itô Isometry

The stochastic term is $I_t = \int_0^t \sigma e^{-\theta(t-u)} dW_u$. Since the integrand is deterministic, $\mathbb{E}[I_t] = 0$. By **Itô Isometry**, $\mathbb{E}[(\int g dW)^2] = \mathbb{E}[\int g^2 du]$:

$$\mathrm{Var}(I_t) = \sigma^2 \int_0^t (e^{-\theta(t-u)})^2 du = \sigma^2 \int_0^t e^{-2\theta(t-u)} du$$

Let $v = t - u$, $du = -dv$. Then:

$$\mathrm{Var}(I_t) = \sigma^2 \int_t^0 e^{-2\theta v}(-dv) = \frac{\sigma^2}{2\theta}(1 - e^{-2\theta t})$$

3 Exact Discretization for State-Space

Most practitioners use Euler approximation ($F \approx 1 - \theta \Delta t$). We derive the **Exact Discretization** to maintain numerical stability.

Definition 3.1: Discrete State Transition

From the continuous solution over a step Δt, we derive the AR(1) form:

$$\epsilon_t = F\epsilon_{t-1} + (1 - F)\mu + w_t$$

where discrete parameters are rigorously linked to physical parameters:

- **Transition Matrix:** $F = e^{-\theta \Delta t}$
- **Process Noise Variance:** $Q = \frac{\sigma^2}{2\theta}(1 - e^{-2\theta \Delta t})$

4 Bayesian Derivation of the Kalman Filter

We reject the algebraic "Minimum Mean Square Error" derivation in favor of the **Bayesian Inference** approach using Joint Gaussian distributions.

4.1 Joint Distribution Construction

Let the prior state be $x \sim \mathcal{N}(\bar{x}, P)$ and measurement $y = Hx + v$ with $v \sim \mathcal{N}(0, R)$. The joint vector $z = [x^T, y^T]^T$ follows:

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \bar{x} \\ H\bar{x} \end{pmatrix}, \begin{pmatrix} P & PH^T \\ HP & HPH^T + R \end{pmatrix} \right) \tag{5}$$

4.2 Conditioning via Schur Complement

Lemma 4.1: Conditional Gaussian Density

Given a joint Gaussian partition $z \sim \mathcal{N}(\mu, \Sigma)$, the conditional density $p(x|y)$ is Gaussian with moments:

$$\mathbb{E}[x|y] = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y) \tag{6}$$

$$\mathrm{Var}(x|y) = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} \tag{7}$$

Proof of Bayesian Update

Define the innovation vector $e = y - (Hx + v)$. Consider the transformation $z' = x - \Sigma_{xy}\Sigma_{yy}^{-1} y$. We compute covariance:

$$\mathrm{Cov}(z', y) = \mathrm{Cov}(x, y) - \Sigma_{xy}\Sigma_{yy}^{-1}\mathrm{Var}(y) = \Sigma_{xy} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yy} = 0$$

Since z' and y are uncorrelated and Jointly Gaussian, they are independent. Thus $\mathbb{E}[x|y] = \mathbb{E}[z'+Ky|y] = \mathbb{E}[z'] + Ky = (\mu_x - K\mu_y) + Ky$. This confirms the Kalman Gain structure $K = \Sigma_{xy}\Sigma_{yy}^{-1}$.

4.3 The Kalman Update Equations

Substituting the sub-matrices from Section 4.1 into the Lemma:

Theorem 4.1: Optimal Posterior Estimates

1. **Kalman Gain:**
$$K = PH^T(HPH^T + R)^{-1}$$
2. **State Update (Posterior Mean):**
$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K(y_t - H\hat{x}_{t|t-1})$$
3. **Covariance Update (Posterior Variance):**
$$P_{t|t} = (I - KH)P_{t|t-1}$$

5 Maximum Likelihood Calibration

To estimate $\Psi = \{\theta, \sigma, R\}$, we maximize the Log-Likelihood of the innovation sequence e_t.

Definition 5.1: Objective Function

$$\mathcal{L}(\Psi) = -\frac{1}{2}\sum_{t=1}^{T}\left(\ln|S_t| + e_t^T S_t^{-1} e_t\right) \tag{8}$$

where $S_t = HP_{t|t-1}H^T + R$ is the innovation covariance.

6 Application to Statistical Arbitrage

This section maps the general theory to the specific problem of dynamic pairs trading.

6.1 State-Space Mapping

We define the dynamic relationship between Asset Y and Asset X as:

$$y_t = \alpha_t + \beta_t x_t + v_t$$

Definition 6.1: System Matrices for Pairs Trading

To apply the filter derived in Section 4, we define:

- **State Vector:** $\boldsymbol{\theta}_t = [\alpha_t, \beta_t]^T$ (Intercept and Hedge Ratio).
- **Observation Matrix:** $H_t = [1, \quad x_t]$.
- **Transition:** $F = \mathbf{I}_2$ (Random Walk assumption for parameters).

6.2 Signal Generation (Z-Score)

The trading signal is derived from the standardized innovation.

Theorem 6.1: Z-Score Definition

Let $e_t = y_t - (\hat{\alpha}_{t|t-1} + \hat{\beta}_{t|t-1} x_t)$ be the raw innovation. Let $S_t = H_t P_{t|t-1} H_t^T + R$ be the theoretical variance. The Z-Score is defined as:

$$Z_t = \frac{e_t}{\sqrt{S_t}} \sim \mathcal{N}(0, 1) \tag{9}$$

Trading Rules:

- **Long Spread:** Enter when $Z_t < -\delta$ (e.g., -2.0).
- **Short Spread:** Enter when $Z_t > \delta$ (e.g., +2.0).
- **Exit:** Close when $|Z_t| < \gamma$ (e.g., 0.5).