An Ornstein-Uhlenbeck Framework for Pairs Trading

Stefan Rampertshammer

Department of Mathematics and Statistics
the University of Melbourne, Parkville Victoria 3010

s.rampertshammer@ms.unimelb.edu.au

November 21, 2007

Contents

Chapter 1

Introduction

1.1 Pairs Trading

Leading up to the late 1980s, when Pairs Trading was developed, there were many shocks to the economy; the Energy Crises of 1970s, recession in 1982-83 and the junk bond collapse in 1988 in particular. This lead to a climate where investors wished to be hedged against market movements. Pairs Trading was developed as a means of making profitable trades while also keeping market exposure low.
Ehrman [5] defines Pairs Trading as:

> *A nondirectional, relative-value investment strategy that seeks to identify two companies with similar trading characteristics whose equity securities are currently trading at a range outside their historical range. This investment strategy entails buying the undervalued security while short-selling the overvalued security; thereby maintaining market neutrality.*

It is in effect a bet that the securities will return to their historical range and so a bet of the two securities against each other rather than against the market.

Pure Arbitrage refers to a situation where a series of trades can be made such that with probability 1, the payoff is non-negative and the payoff is positive with positive probability, that is a situation where riskless profit is possible. Pairs Trading does not meet these criteria as engaging in the pairs trade exposes the trader to fluctuations in the stock prices, or market risk.

This sort of trade is regarded as Statistical Arbitrage, which only requires that the expected value of the trade be positive. A single Statistical Arbitrage trade has positive probability of loss but the usual context of the trade is where many similar trades are also being made or this trade is repeated over time. As the number of trades becomes large, the Law of Large Numbers implies that the average returns approach their expected value. As this expected value is positive, when we repeat this process many times, an Arbitrage situation arises.

This project is concerned with modeling situations in which a pairs trade is possible and profitable and determining an 'optimal' trade if it exists.

1.2 Factor Models

To attempt any mathematical analysis of Pairs Trading, we first require a mathematical framework for the underlying financial market.

We define the *return* on an asset in some time period to be the percentage change in its value, X, that is

$$r_X = \frac{X_{\text{final}} - X_{\text{initial}}}{X_{\text{initial}}} \approx \log\left(\frac{X_{\text{final}}}{X_{\text{initial}}}\right).$$

This notion makes sense for any asset, but our main concern is with stocks, so from here any asset will assumed to be a tradable stock.

Arbitrage Pricing Theory supposes that a stock's return, r_X, in some time period is made up of two main components: the stock specific component, r_X^e which is some company specific constant, and a linear combination of some factors $f^{(i)}$, $i = 1, 2, \ldots, n$. That is:

$$r_X = r_X^e + \sum_{i=1}^{n} \beta_X^{(i)} f^{(i)} + \epsilon_X$$

$$\mathbb{E}[r_X] = r_X^e + \sum_{i=1}^{n} \beta_X^{(i)} \mathbb{E}[f^{(i)}]$$

where ϵ_X is a random variable with zero mean and uncorrelated with $f^{(i)}$, $\beta_X^{(i)}$ is a measure of the sensitivity to factor i of stock X and is given by

$$\beta_X^{(i)} = \frac{\text{cov}(r_X, f^{(i)})}{\text{var}(f^{(i)})}$$

when the $f^{(i)}$ are uncorrelated.

If we expand the time horizon to include more than one period then we must make some assumptions:

- Assumptions on r_X^e:
  r_X^e has the natural interpretation of being a numerical representation of the company fundamentals; these are things that will not change without a major restructuring of the company so it is reasonable to assume that r_X^e is constant in time, at least in the periods we will be considering. Note that any company specific deviation from $\mathbb{E}[r_X]$ is captured in ϵ_X.

- Assumptions on $\beta_X^{(i)}$:
  $\beta_X^{(i)}$ represents the sensitivity of stock X to factor i. This is also build into the structure of X and so will be constant in the short run.

- Assumptions on $f^{(i)}$:
  These are exogenous factors indicative of the state of the market. For example, the Capital Asset Pricing Model (CAPM) has one factor which is the return on the market portfolio above the riskfree rate. Another popular model, the Fama-French Model has 3 factors. In our use, exogenous means that their values are determined outside the model, we view $(f_t^{(1)}, f_t^{(2)}, \dots f_t^{(n)})$ as a process which we observe and cannot affect.

  We allow relationships between the $f^{(i)}$, but in any sensible model they would not be perfectly correlated. If this were the case, a smaller model would capture exactly the same information and would be preferable. These factors are changing with time, but as mentioned before, are exogenous so that they are unaffected by the values of r_X. As these factors represent fundamental aspects of the market we would not expect the dependence structure between $f^{(i)}$ and $f^{(j)}$ to change over time. So we assume that, or any $s, t, h > 0$,

$$\mathbb{E}\left[f_t^{(i)} f_s^{(j)}\right] = \mathbb{E}\left[f_{t+h}^{(i)} f_{s+h}^{(j)}\right].$$

1.3 Desirable Model Properties

1.3.1 Real World Approach

Presently, Pairs Trading is one of the many techniques used by large financial institutions. As there are so many pairs in the trading universe to consider, it takes a lot of data to find the good candidates for potential pairs trades. By looking at what is done in the real world we can extract the important characteristics we would like in a model approximating this elusive real world.

The steps involved with a (real world) pairs trade, according to Vidyamurthy [16], are as follows:

1. Identify potential stock pairs.
2. Verify that a pairs trade is appropriate.
3. Determine the ratio of stocks to buy and short sell.

Identify Potential Stock Pairs

The feature we would most like to have in a pair is *co-integration*, loosely speaking this guarantees that the stock price will return to some known level in a reasonable time, a more precise definition will follow. A necessary condition for co-integration between some series is that correlation between the series be 1, but this is extremely rare so we compromise and take all pairs that are almost perfectly correlated as our candidates for a pairs trade. Vidyamurthy [16] goes into more detail about what constitutes 'almost' in this sense. Note that we are concerned with price ratios so is convenient in this analysis to consider log-prices.

Verify If A Pairs Trade is Appropriate

(Almost) Perfect correlation gives a necessary condition for co-integration but not a sufficient one. With a potential pair we do the following:

1. Determine the linear combination.
   We have a pair of highly correlated stocks, that means that there exists some linear relationship between them. To find it, we run a regression of one stock's returns against the other's.

2. Test for co-integration.

   A desirable property of a pair (for pairs trading) is that the ratio of prices returns to some historical level, we refer to this as mean reversion. Given our linear relationship between the two log-price series, we test that what is left over returns to 0 sufficiently often.

If the residuals are sufficiently mean reverting then a Pairs Trade is appropriate for the pair of stocks.

Determine Details of the Trade

The slope of the linear combination determined above will be the ratio a which to trade. We do this so that our returns maintain the market neutrality we desire. For example, if the linear combination between $x = \log(X)$ and $y = \log(Y)$ were discovered to be $y = \gamma x$ then we would trade $x : y$ in the ratio $\gamma : 1$.

These are the properties we would like to mimic in a formulated model.

Now to properly define some of these terms:

1.3.2 Stationarity

Definition 1. (Strict Stationarity)
A process X_t is *stationary in the strict sense* if, for any $\{t_1, \ldots, t_k\}$ where $k = 1, 2, \ldots;$

$$F_{t_1,\ldots,t_k} = F_{t_1+h,\ldots,t_k+h}$$

for all h, where $F_{t_1,\ldots,t_k}$ is the distribution function of $(X_{t_1}, \ldots, X_{t_k})$.

Definition 2. (Weak Stationarity)
A square integrable process X_t is *stationary in the weak (wide) sense* if

$$\mathbb{E}(X_t) = \mu(t) = \mu$$

is $t-$invariant and

$$\text{cov}(X_t, X_s) = f(|t - s|).$$

This implies that $\text{var}(X_t) = \text{cov}(X_t, X_t) = f(0)$ is also constant in t.

As a Gaussian process is completely specified by its mean and covariance functions, for this class of processes, weak and strong stationarity are equivalent.

1.3.3 Co-Integration

Definition 3. (Co-Integration)
Two nonstationary time series A_t, B_t are *co-integrated* if there is some $\gamma \in \mathbb{R}$ such that $A_t - \gamma B_t$ is stationary in the weak sense.

If we put co-integration into a factor model framework, co-integration of asset returns of X and Y is equivalent to

$$\beta_X^{(i)} = \gamma \beta_Y^{(i)} \tag{1.1}$$

for all i.

First assuming that the relationship in 1.1 holds:
By the factor model:

$$r_{X,t} = r_X^e + \sum_{i=1}^{n} \beta_X^{(i)} f_t^{(i)} + \epsilon_{X,t},$$
$$r_{Y,t} = r_Y^e + \sum_{i=1}^{n} \beta_Y^{(i)} f_t^{(i)} + \epsilon_{Y,t},$$

where the $\epsilon_{\cdot,t}$ are independent and have zero mean.

So if, for some γ, $\beta_X^{(i)} = \gamma \beta_Y^{(i)}$ for all i then

$$\mathbb{E}(r_{X,t} - \gamma r_{Y,t}) = r_X^e - \gamma r_Y^e,$$

which is constant over time. Also

$$\mathbb{E}\left[(r_{X,t} - r_X^e)(r_{Y,s} - r_y^e)\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} \beta_X^{(i)} \beta_Y^{(j)} \mathbb{E}[f_t^{(i)} f_s^{(j)}],$$

which depends only on $|t-s|$ by assumptions on $f_t^{(i)}$, so the returns are co-integrated with factor of co-integration γ.

Now if the series are co-integrated, it remains for us to show that $\beta_X^{(i)} = \gamma \beta_Y^{(i)}$ for all i. But this is not difficult as

$$\mathbb{E}(r_{X,t} - \gamma r_{Y,t}) = r_X^e - \gamma r_Y^e + \sum_{j=1}^{n} (\beta_X^{(i)} - \gamma \beta_Y^{(i)}) \mathbb{E}[f_t^{(i)}]$$

so, as $\mathbb{E}[f_t^{(i)}]$ changes over time, the only way that $\mathbb{E}(r_{X,t} - \gamma r_{Y,t})$ can be constant over time is if $\beta_X^{(i)} = \gamma \beta_Y^{(i)}$ for all i.

Formal testing of co-integration of real world time series has some difficulties, some of which are detailed in [13]. We are more concerned with developing a model which has co-integration as a feature than determining co-integration on existing series though.

1.3.4 Ergodicity

The Pairs Trade relies on Statistical Arbitrage. We are betting that the price ratio will return to some historical value, so in our model we will require that, with probability 1, we will return to our historical value. Moreover, we would like to return there in finite time.

Definition 4. (Ergodicity) A $d-$dimensional random process, $\{\mathbf{X}_t, t \geq 0; \mathbf{X}_0 = \mathbf{x}\}$ is *ergodic* if there is some measure μ such that, for any $\mu-$integrable function f,

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t f(\mathbf{X}_s) ds = \int_{\mathbb{R}^d} f d\mu.$$

This means that the long run average behavior mimics the instantaneous behavior. So an ergodic process will visit each state infinitely often, and this implies that the time between visits is certainly finite.

Theorem 1. [11] (Hasminski):
For a diffusion process, $\{\mathbf{X}_t\}_{t \geq 0}$ with stochastic differential equation $d\mathbf{X}_t = \mu(\mathbf{x})dt + \sigma(\mathbf{x})d\mathbf{W}_t$, $\mu(\mathbf{x}) \in \mathbb{R}^d$ and $\sigma(\mathbf{x}) \in \mathbb{R}^{d \times d}$, the following are sufficient conditions for ergodicity:

- $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$ are smooth with bounded derivatives of any order and $\sigma(\mathbf{x})$ is additionally bounded, that is each element of $\sigma(\mathbf{x})$ is bounded.
- There exists $\beta > 0$ such that, for some compact subset $K \subset \mathbb{R}^d$,

  $$\mathbf{x}^T \mu(\mathbf{x}) \leq -\beta \mathbf{x}^T \mathbf{x}$$

  for all $\mathbf{x} \in \mathbb{R}^d \setminus K$, where $\mathbf{x}^T$ denotes the transpose of $\mathbf{x}$.

Tighter conditions for ergodicity exist, but these conditions will suffice for the models considered here.

1.4 Stochastic Integration

1.4.1 Itô Integral

The exposition in this subsection comes from [10]:
We wish to make sense of

$$\int_0^t b_s d\mathbf{W}_s,$$

where $\{\mathbf{W}_t\}_{t\geq 0}$ is the standard Brownian motion and $\{b_t\}_{t\geq 0}$ is a process adapted to $\{\mathbf{W}_t\}$.

The Itô Integral is defined for simple processes $b_t = \sum_{k=0}^{n} c_k \mathbb{1}_{(t_{k-1},t_k]}$ as

$$\int_0^t b_s dW_s = \sum_{k=1}^{n-1} c_k \left(\mathbf{W}_{t_{k+1}} - \mathbf{W}_{t_k}\right)$$

Note that if the c_k are nonrandom then $\int_0^t b_s d\mathbf{W}_s$ is necessarily Gaussian.

Then, taking c_k as $\mathcal{F}_{t_k}-$measurable with finite variance to consider a more general class of functions that need not be Gaussian. We then take the standard measure theory approach of an approximating sequence of simple functions to extend the definition of the Itô integral to a wider class of processes.

A couple of relevant properties of the integral defined this way are:

$$\mathbb{E}\left[\int_0^t b_s d\mathbf{W}_s\right] = 0 \tag{1.2}$$

$$\mathbb{E}\left[\left(\int_0^t b_s d\mathbf{W}_s\right)^2\right] = \int_0^t \mathbb{E}[b_s^2] ds \tag{1.3}$$

So combining these observations gives that, if $b_s = f(s)$ is nonrandom

$$\int_0^t f(s) d\mathbf{W}_s \sim \mathcal{N}\left(0, \int_0^t f(s)^2 ds\right),$$

where $\mathcal{N}(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2.

Definition 5. We call $\{\mathbf{X}_t\}$ an *Itô process* if

$$\mathbf{X}_t = \mathbf{X}_0 + \int_0^t a_s ds + \int_0^t b_s d\mathbf{W}_s$$

for some $\{a_t\}, \{b_t\}$ adapted to $\{\mathbf{W}_t\}$.
In this case we say $\mathbf{X}_t$ has a *stochastic differential* given by

$$d\mathbf{X}_t = a_t dt + b_t d\mathbf{W}_t.$$

1.4.2 Diffusion Processes

Definition 6. We call $\{\mathbf{X}_t\}$ a diffusion process if $\mathbf{X}_t$ is an Itô process and Markovian, that is, for $\{\mathcal{F}_t\}$ the $\sigma-$algebra generated by the process up to t, $\mathbb{E}\left[g(\mathbf{X}_{t+h})|\mathcal{F}_t\right] = \mathbb{E}\left[g(\mathbf{X}_{t+h})|\mathbf{X}_t\right]$ for all $h > 0$ and all bounded measurable g.

Any process with a stochastic differential given by

$$d\mathbf{X}_t = a(t, \mathbf{X}_t)dt + b(t, \mathbf{X}_t)d\mathbf{W}_t$$

is a diffusion process. The evolution of the process from time t only requires knowledge of the process at time t, and not its prior progress so that the Markovian property is satisfied.

1.4.3 Linear SDEs

The exposition in this subsection uses material pursued in [8], Chapter 5:
A linear SDE is one of the form:

$$d\mathbf{X}_t = \left[A(t)\mathbf{X}_t + a(t)\right] dt + \sigma(t)d\mathbf{W}_t, \qquad \mathbf{X}_0 = \mathbf{X}_0, \tag{1.4}$$

where $\mathbf{X}_t \in \mathbb{R}^d$.
It has an associated deterministic differential equation

$$d'(t) = A(t)d(t) + a(t), \qquad d(0) = \mathbf{X}_0. \tag{1.5}$$

Standard tenets of calculus guarantee a unique, absolutely continuous solution, d, to this initial value problem.

Consider $\Phi : \mathbb{R} \to \mathbb{R}^d$ satisfying

$$\Phi'(t) = A(t)\Phi(t), \qquad \Phi(0) = I.$$

Then the solution to the deterministic part of (1.4) is

$$d(t) = \Phi(t)\left[d(0) + \int_0^t \Phi^{-1}(s)a(s)ds\right].$$

Note that $\Phi^{-1}(t)$ is guaranteed to exist otherwise there would be some t_0 and $v \neq 0$ such that $\Phi(t_0)v = 0$ which would imply that $\Phi(t_0)v$ is a solution to $d'(t) = A(t)d(t)$. But our solution is unique so that this would imply $\Phi(t)v = 0$ for all t which contradicts $\Phi(0) = I$. Thus $\Phi(t)$ is nonsingular.

Then, by Itô's formula the process given by

$$\mathbf{X}_t = \Phi(t)\left[\mathbf{X}_0 + \int_0^t \Phi^{-1}(s)a(s)ds + \int_0^t \Phi^{-1}(s)\sigma(s)dW_s\right]$$

satisfies (1.4).

Indeed,

$$\begin{aligned} dX_t &= A(t)X_t dt + \Phi(t)\left[\Phi^{-1}(t)a(t)dt + \Phi^{-1}(t)\sigma(t)dW_t\right] \\ &= \left[A(t)X_t + a(t)\right]dt + \sigma(t)dW_t. \end{aligned}$$

1.4.4 Kolmogorov Backward Equations

Let us consider the process $\{\mathbf{X}_t\}$, $\mathbf{X}_t \in \mathbb{R}$ with stochastic differential $d\mathbf{X}_t = a(t, \mathbf{X}_t)dt + b(t, \mathbf{X}_t)d\mathbf{W}_t$, $\mathbf{W}_t \in \mathbb{R}^d$ a vector whose components are independent brownian motions. We term

$$a(t, \mathbf{x}) = \left(a^{(1)}(t, \mathbf{x}), \ldots, a^{(d)}(t, \mathbf{x})\right)^T$$

the *Drift Vector* and

$$b(t, \mathbf{x}) = \left[b^{(i,j)}(t, \mathbf{x})\right], \quad i, j = 1, 2, \ldots, d,$$

the *Diffusion Matrix*.

As in [11], Chapter 2. Let $u(s, x)$ be some function differentible in s and twice differentible in $\mathbf{x}$. Define the operator $\mathcal{A}$ by

$$\mathcal{A}u(s, \mathbf{x}) = \sum_{i=1}^{d} a^{(i)}(s, \mathbf{x})u_{x_i}(s, \mathbf{x}) + \frac{1}{2}\sum_{i,j=1}^{d} b^{(i,j)}(s, \mathbf{x})u_{x_i x_j}(s, \mathbf{x}),$$

where $u_{x_i}(s, \mathbf{x}) = \frac{\partial}{\partial x_i}u(s, \mathbf{x})$.

Definition 7. The *Kolmogorov Backwards Equation* for the process $\{\mathbf{X}_t\}$ is given by:

$$u_s(s, \mathbf{x}) + \mathcal{A}u(s, \mathbf{x}) = 0. \tag{1.6}$$

Claim 1. Let, for free $t > 0$, $y \in \mathbb{R}$, $u(s, \mathbf{x}) = p(s, \mathbf{x}; t, \mathbf{y})$ is the transition density from $\mathbf{X}_s = \mathbf{x}$ to $\mathbf{X}_t = \mathbf{y}$ along with the condition that $\lim_{s \uparrow t} u(s, \mathbf{x}) = \delta(\mathbf{x} - \mathbf{y})$ for δ the Dirac Delta-function on $\mathbb{R}^d$. Then $u(s, \mathbf{x})$ is a solution to (1.6).

Claim 2. Let $u(s, \mathbf{x}) = \mathbb{E}\left[\psi(t, \mathbf{X}_t) | \mathbf{X}_s = \mathbf{x}\right]$ with the condition that $\lim_{s \uparrow t} u(s, \mathbf{x}) = \psi(t, x)$. Then $u(s, \mathbf{x})$ is a solution to (1.6).

For proof of Claim 1 see [7], Chapter 12.

Proof of Claim 2:
By Claim 1, $v(s, \mathbf{x}) = p(s, \mathbf{x}; t, \mathbf{y})$ solves $v_s(s, \mathbf{x}) + \mathcal{A}v(s, \mathbf{x}) = 0$ for all $t > s$ and $\mathbf{y} \in \mathbb{R}^d$. Then

$$\begin{aligned} u(s, \mathbf{x}) &= \mathbb{E}\left[\psi(t, \mathbf{X}_t) | \mathbf{X}_s = \mathbf{x}\right] \\ &= \int_{\mathbb{R}^d} p(s, \mathbf{x}; t, \mathbf{y}) \psi(t, \mathbf{y}) d\mathbf{y}, \end{aligned}$$

and so u is a solution to $u_s(s, \mathbf{x}) + \mathcal{A}u(s, \mathbf{x}) = 0$ by linearity of $\mathcal{A}$.

As s and t become close $\lim_{s \uparrow t} v(s, \mathbf{x}) = \delta(\mathbf{x} - \mathbf{y})$ so

$$\begin{aligned} \lim_{s \uparrow t} u(s, \mathbf{x}) &= \int_{\mathbb{R}^d} \delta(\mathbf{x} - \mathbf{y}) \psi(s, \mathbf{y}) d\mathbf{y} \\ &= \psi(s, \mathbf{x}). \end{aligned}$$

Chapter 2

Ornstein-Uhlenbeck Processes

2.1 Models Considered

We will be specifically considering the following two models:

1. One Process Model

$$X_t = x \exp\{F_t\} \tag{2.1}$$

$$Y_t = kX_t \exp\{\tilde{Y}_t\}, \quad t \geq 0, \tag{2.2}$$

where $\{F_t\}$ is a factor component and $\{\tilde{Y}_t\}$ is an Ornstein-Uhlenbeck process which may or may not depend on F_t.k is a multiplicative constant which reflects the historical ratio of stock prices. x represents the price of X in the absence of other factors.

2. Two Processes Model

$$X_t = x \exp\{F_t + \tilde{X}_t\} \tag{2.3}$$

$$Y_t = y \exp\{F_t + \tilde{Y}_t\}, \quad t \geq 0, \tag{2.4}$$

where $\{F_t\}$ is a common factor component and $\{(\tilde{X}_t, \tilde{Y}_t)\}$ is a two dimensional Ornstein-Uhlenbeck process. We will take $y = kx$ where k is representative of the historical price ratio between stocks X and Y. x and y represents the prices of X and Y, respectively, in the absence of other factors.

In both cases we consider the factor component F_t to be a linear combination of factors $F_t^{(i)}$ such that X and Y have the same sensitivities to all factors. This

implies that any departure from the historical price ratio is due to the movements of $\tilde{Y}_t$ (and $\tilde{X}_t$ for the second model).

The Ornstein-Uhlenbeck process arises naturally here as it is a mean reverting process; the drift vector is giving the process a tendency to return to zero (equilibrium)

2.1.1 One Process Model

Recall (2.1). This is a model that supposes that stock X is fairly stable and can be well predicted through factors $\{F_t\}$ and stock Y fluctuates around it.

In equilibrium ($\tilde{Y}_t = 0$), the stock prices are in ratio ($Y : X$) $1 : k$, so if we are to maintain a market neutral position, a natural trade to make would be in this ratio. We trade when the ratio of stock prices becomes sufficiently different from $1 : k$.

Note that if $\frac{Y_t}{kX_t} = a$ then $\log(Y_t) - \log(kX_t) = \log(a)$, so dealing with log differences is equivalent to dealing with ratios. We choose to use the former.

Note also that $\log(Y_t) - \log(kX_t) > 0$ means that in some sense Y_t is overvalued relative to X_t so the correct thing to do would be to short Y and long X, and vice versa when $\log(Y_t) - \log(kX_t) < 0$.

Let $b_1 < 1$, $b_2 > 1$ and τ be the time that $\log(Y_t) - \log(kX_t) = 0$ after $\log(Y_t) - \log(kX_t) = \log(b_i)$, $i = 1, 2$.

- When $\log(Y_t) - \log(kX_t) = \log(b_1) < 0$ we short k units of stock X and long 1 unit of stock Y and reverse at time τ.
- When $\log(Y_t) - \log(kX_t) = \log(b_2) > 0$ we long k units of stock X and short 1 unit of stock Y and reverse at time τ.

If we consider for a moment the b_1 case, the net transaction at the initial trade, time T is

$$kX_T - Y_T = Y_T(b_1 - 1) = kX_T(1 - b_1^{-1}).$$

At the conclusion of the trade, $-kX_\tau + Y_\tau = 0$.

This leaves a net return of

$$Y_T(b_1 - 1) = kX_T(1 - b_1^{-1}).$$

Similarly it can be shown the second case yields a return of

$$Y_T(1 - b_2) = kX_T(b_2^{-1} - 1).$$

Note that both of these are necessarily positive.

This section ignores the time value of money. This is done because we are dealing with a short term investment so that the change in value over time is negligible. We have also assumed that the transaction costs of a single trade are negligible so that transaction costs only become relevant when many trades are made in a short time.

2.1.2 Two Processes Model

Recall (2.3). In this model both stocks have a common factor component $\{F_t\}$ so all of the departure from the historical ratio is captured in $(\tilde{X}_t, \tilde{Y}_t)$. The dependence of $\tilde{X}_t$ and $\tilde{Y}_t$ plays a large part in how the stock prices evolve jointly.

We suppose that, without any deviations from the mean value, the stocks are in a $1 : k$ price ratio. We write $y = kx$, so that k is indicative of the historical price ratio of stocks X and Y. Then we adopt the same trading strategy, namely, for b_1, b_2, τ defined as before.

- When $\log(Y_t) - \log(kX_t) = \log(b_1) < 0$ we short k units of stock X and long 1 unit of stock Y and reverse at τ.
- When $\log(Y_t) - \log(kX_t) = \log(b_2) > 0$ we long k units of stock X and short 1 unit of stock Y and reverse at τ.

Also as above, the net returns are given by:

$$\begin{aligned} Y_T(b_1 - 1) &= kX_T(1 - b_1^{-1}), \qquad b_1 > 1, \\ Y_T(1 - b_2) &= kX_T(b_2^{-1} - 1), \qquad b_2 < 1. \end{aligned}$$

2.2 Functions of Matrices

The exposition in this section is based on the material presented in Chapter 5 of [12].

For a polynomial $p(z) = \sum_{j=0}^{n} a_j z^j$ and $A \in \mathbb{C}_{n\times n}$, the set of all $n \times n$ matrices with complex coefficients, define

$$p(A) = \sum_{j=0}^{n} a_j A^j.$$

Let $A \in \mathbb{C}_{n\times n}$ have minimal polynomial $\psi(z)$,

$$\psi(z) = (z-\lambda_1)^{n_1}(z-\lambda_2)^{n_2}\dots(z-\lambda_m)^{n_m},$$

where λ_i are distinct eigenvalues of A.

We say that two functions, g and h agree on the *spectrum* of A if, for all $i = 1, 2, \dots, m$ and $j = 0, 1, \dots, n_i$,

$$\frac{d^j}{dz^j} g(z)\big|_{z=\lambda_i} = \frac{d^j}{dz^j} h(z)\big|_{z=\lambda_i}.$$

Given ψ then for any polynomial p there are polynomials q and r such that r is of lesser degree than ψ and

$$p(z) = q(z)\psi(z) + r(z).$$

Recall that $\psi(A) = 0$ by definition of the minimal polynomial. It follows that $p(A) = r(A)$

Lemma 1. For g, h polynomials and $A \in \mathbb{C}_{n\times n}$ with eigenvalues $\lambda_1, \dots, \lambda_m$; $g(A) = h(A)$ iff g agrees with h on the spectrum of A.

Theorem 2. If f is defined on the spectrum of A and g is the polynomial of minimum degree determined by the values of f on the spectrum of A then $f(A)$ exists and $f(A) := g(A)$.

Theorem 3. Let $A \in \mathbb{C}_{n\times n}$ have eigenvalues $\lambda_1, \dots, \lambda_m$ and let f have Taylor Series around z_0 given by

$$f(z) = \sum_{k=0}^{\infty} \alpha_k (z - z_0)^k,$$

converging in radius r such that $|\lambda_j - z_0| < r$ for all $\lambda_1, \ldots, \lambda_m$. Then $f(A)$ exists and is given by

$$f(A) = \sum_{k=0}^{\infty} \alpha_k (A - z_0 I)^k$$

.

This means in particular that

$$e^A = \sum_{r=0}^{\infty} \frac{1}{r!} A^r$$

for all $A \in \mathbb{C}_{n\times n}$

Theorem 4. Let $G(z_1, \ldots, z_l)$ be a polynomial in $z_1, \ldots, z_l$ and $f_i, i = 1, 2, \ldots, l$ be functions defined on the spectrum of $A \in \mathbb{C}_{n\times n}$ such that $G(f_1, f_2, \ldots, f_l) = 0$ on the spectrum of A then

$$G(f_1(A), f_2(A), \ldots, f_l(A)) = 0.$$

The proofs of all these theorems use Hermite polynomial interpolation on the spectrum of A and then Lemma 1.

A consequence of this is that if A and B commute, then $e^{A+B} = e^A e^B$.

2.3 Univariate Process

Let $\{W_t\}$ be a standard brownian motion. A stochastic process $\{X_t : t \geq 0\}$ is called an *Ornstein-Uhlenbeck* (OU) process if it satisfies the following stochastic differential equation for constants ρ, σ, θ:

$$dX_t = -\rho(X_t - \theta)dt + \sigma dW_t.$$

Note that if $\{X_t\}$ is OU then, for any $\alpha \in \mathbb{R}$, $\{X_t - \alpha\}$ is also OU, so without loss of generality, we can let $\theta = 0$.

This is a diffusion process and so is automatically Markovian.

2.3.1 Explicit Solution

Using that this is a linear stochastic differential equation, and following section 1.4.3, we solve the deterministic part of the equation which yields

$$X_t = Ce^{-\rho t}.$$

If we consider $C = C_t$ as a process adapted to $\{W_t\}$ then we obtain

$$X_t = e^{-\rho t}\left[X_0 + \int_0^t \sigma e^{\rho r} dW_r\right]. \tag{2.5}$$

The integrand in the Itô integral $\int_0^t \sigma e^{\rho r} dW_r$ is nonrandom, so that the process has independent increments (in particular, is independent of X_0) and hasa normal distribution

$$\int_0^t \sigma e^{\rho r} dW_r \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\rho}\left(e^{2\rho t} - 1\right)\right).$$

This means

$$\mathbb{E}[X_t] = \mathbb{E}[X_0]e^{-\rho t}, \tag{2.6}$$

$$\operatorname{cov}(X_t, X_s) = e^{-\rho(t+s)}\left(\operatorname{var}(X_0) + \frac{\sigma^2}{2\rho}\left(e^{2\rho(t\wedge s)} - 1\right)\right). \tag{2.7}$$

Indeed,

$\mathbb{E}[X_t]$ is clearly seen from taking expectations of (2.5).

To prove (2.7), we write

$$\begin{aligned}\operatorname{cov}(X_t, X_s) &= \mathbb{E}\left[\left(X_t - e^{-\rho t}\mathbb{E}[X_0]\right)\left(X_s - e^{-\rho s}\mathbb{E}[X_0]\right)\right], \\ &= \mathbb{E}[X_t X_s] - \mathbb{E}[X_0]^2 e^{-\rho(t+s)},\end{aligned}$$

where

$$\begin{aligned}\mathbb{E}[X_t X_s] &= \mathbb{E}\left[\left(e^{-\rho t}\left[X_0 + \int_0^t \sigma e^{\rho r} dW_r\right]\right)\left(e^{-\rho s}\left[X_0 + \int_0^s \sigma e^{\rho r} dW_r\right]\right)\right] \\ &= \mathbb{E}\left[e^{-\rho(t+s)}\left(X_0^2 + \sigma^2 \int_0^t e^{\rho v} dW_v \int_0^s e^{\rho r} dW_r\right)\right]\end{aligned}$$

Then, as increments are independent, only the overlapping intervals add to the expectation of the integral. So by this, (1.2) and (1.3)

$$\begin{aligned}\mathbb{E}\left[\int_0^t e^{\rho v} dW_v \int_0^s e^{\rho r} dW_r\right] &= \mathbb{E}\left[\left(\int_0^{t\wedge s} e^{\rho r} dW_r\right)^2\right] \\ &= \int_0^{t\wedge s} e^{2\rho r} dr \\ &= \frac{1}{2\rho}(e^{\rho(t\wedge s)} - 1),\end{aligned}$$

where $t \wedge s$ is the minimum of s and t.

So the covariance is given by $e^{-\rho(t+s)}\left(\mathbb{E}[X_0^2] - \mathbb{E}[X_0]^2 + \frac{\sigma^2}{2\rho}\left(e^{2\rho(t\wedge s)} - 1\right)\right)$ and we are done.

If $X_0 = x$ with probability 1, then X_t has distribution

$$(X_t | X_0 = x) \sim \mathcal{N}\left(xe^{-\rho t}, \frac{\sigma^2}{2\rho}\left(1 - e^{-2\rho t}\right)\right) \tag{2.8}$$

and has mean and covariance functions given by (2.7) and (2.6) by:

$$\begin{aligned}\mathbb{E}[X_t | X_0 = x] &= xe^{-\rho t} \\ \operatorname{cov}(X_t, X_s | X_0 = x) &= \frac{\sigma^2}{2\rho}\left(e^{-\rho|t-s|} - e^{-\rho(t+s)}\right).\end{aligned}$$

2.3.2 Stationary Distribution

We now consider the possibility of a stationary distribution for X_t. When $X_0 = x$ is given we see from (2.8) that (when $\rho > 0$), letting $t \to \infty$, yields a

limiting distribution given by

$$X_\infty \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\rho}\right).$$

This is irrespective of the initial value $X_0 = x$ and so is a candidate for the stationary distribution.

If we consider $X_0 \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\rho}\right)$, as

$$\int_0^t \sigma e^{\rho r} dW_r \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\rho}\left(e^{2\rho t} - 1\right)\right)$$

and it is independent of X_0, it is easily seen that for all $t \geq 0$,

$$X_t \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\rho}\right).$$

Moreover, by (2.7)

$$\operatorname{cov}(X_t, X_s) = \frac{\sigma^2}{2\rho} e^{-\rho|t-s|}.$$

Then, as $\mathbb{E}(X_t) = 0$ and $\operatorname{cov}(X_t, X_s) = f(|t-s|)$, $\{X_t\}$ stationary in the weak sense.

It is also stationary in the strong sense. From [11]: If we impose a time shift on (2.8), then the transition density from $X_s = x$ to $X_t = y$, $(t > s)$ is given by

$$p(s, x; t, y) = \frac{\sqrt{\rho}}{\sqrt{\pi\sigma^2(1 - e^{-2\rho(t-s)})}} \exp\left\{-\frac{\rho(y - xe^{-\rho(t-s)})^2}{\sigma^2(1 - e^{-2\rho(t-s)})}\right\}$$

This depends on t and s through $(t-s)$ only, so we can re-express $p(s, x; t, y)$ as $p(x; t-s, y)$.
Then take $F_{t_1,t_2,\ldots,t_k}(x_1, x_2, \ldots, x_k) = \mathbb{P}(X_{t_1} \leq x_1, \ldots, X_{t_k} \leq x_k)$. By independent increments,

$$F_{t_1,t_2,\ldots,t_k}(x_1, x_2, \ldots, x_k) = F_{t_1}(x_1) F_{t_2-t_1}(x_2 - x_1) \ldots F_{t_k-t_{k-1}}(x_k - x_{k-1}).$$

If we then shift time by h, nothing changes. Indeed, let $t_i^* = t_i + h$

$$\begin{aligned} F_{t_i^* - t_{i-1}^*}(x_i - x_{i-1}) &= F_{(t_i+h)-(t_{i-1}+h)}(x_i - x_{i-1}) \\ &= F_{t_i - t_{i-1}}(x_i - x_{i-1}). \end{aligned}$$

Also,

$$F_{t_1^*} = F_{t_1+h}(x_1) = F_{t_1}(x_1).$$

Then, as

$$F_{t_1^*,t_2^*,\ldots,t_k^*}(x_1, x_2, \ldots, x_k) = F_{t_1^*}(x_1)F_{t_2^*-t_1^*}(x_2 - x_1)\ldots F_{t_k^*-t_{k-1}^*}(x_k - x_{k-1}),$$

strong stationarity is assured.

2.3.3 Ergodicity

Recalling the sufficient conditions for ergodicity from Theorem 1, Chapter 1, we need to check the following:

- $-\rho x$ and σ are smooth with bounded derivatives of any order and $\sigma(x)$ is bounded.

  This is true by inspection

- There exists a $\beta > 0$ such that, for some compact subset $K \subset \mathbb{R}$,

  $$-\rho x^2 \leq -\beta x^2$$

  for all $x \in \mathbb{R} \setminus K$

  If $\rho > 0$ then this is satisfied when we take $\beta = \rho$.

Thus when $\rho > 0$ we are assured ergodicity.

When $\rho = 0$ we are left with a pure Brownian motion and it is well known that the time taken to reach any level (different from a fixed starting point) has infinite expectation and so is not at all suitable for the model for which we are considering it.
When $\rho < 0$ the process is repelled from 0. The model we consider requires that the process returns to 0 but in this instance the process is actively discouraged from doing so, thus it is inappropriate for the model we consider.

Thus we only consider processes where $\rho > 0$ and so our process is ergodic.

2.4 Multivariate Process

Let $\{\mathbf{W}_t\}$ be a d-dimensional standard Brownian motion; that is the components of $\{\mathbf{W}_t\}$ are independent univariate standard Brownian Motions. A process $\{\mathbf{X}_t\}$ is called an Ornstein-Uhlenbeck process if, for some constant $d \times d$ matrices A, Σ with Σ positive definite, that is Σ is symmetric and has only positive eigenvalues, one has

$$d\mathbf{X}_t = -A\mathbf{X}_t dt + \Sigma d\mathbf{W}_t.$$

Note that we can be more expansive here and include a vector to which the process is attracted, we use $\mathbf{0}$ without loss of generality.

2.4.1 Solution

We solve this multivariate SDE by simultaneously solving the one-dimensional equations that arise from considering components separately. As each of these is a linear SDE, we can use the approach in Section 1.4.3. Solving the deterministic part gives

$$\mathbf{X} = e^{-At}\mathbf{C}.$$

Next, considering $\mathbf{C} = \mathbf{C}_t$ as a process adapted to $\{\mathbf{W}_t\}$, gives us that

$$\mathbf{X}_t = e^{-At}\left(\mathbf{X}_0 + \int_0^t e^{Ar}\Sigma d\mathbf{W}_r\right). \tag{2.9}$$

As this is a diffusion process, $\Sigma\Sigma^T$ is a matrix of instantaneous covariances and so must be positive definite. For a general Σ, let $\overline{\Sigma} = (\Sigma\Sigma^T)^{1/2}$. It is clear that $\overline{\Sigma}$ is positive definite, we see that $\overline{\Sigma}\overline{\Sigma}^T = \Sigma\Sigma^T$ and so without loss of generality, we can take Σ to be positive definite.

2.4.2 Ergodicity

Recalling the sufficient conditions for ergodicity from Theorem 1 of Chapter 1, we need to check the following:

- $-A\mathbf{x}$ and Σ are smooth with bounded derivatives of any order and Σ is bounded.

  This is true by inspection

- There exists $\beta > 0$ such that, for some compact subset $K \subset \mathbb{R}^d$,

$$-\mathbf{x}^T A\mathbf{x} \leq -\beta||\mathbf{x}||^2$$

for all $\mathbf{x} \in \mathbb{R}^d \setminus K$.

Claim 3. This is true when $\frac{1}{2}(A^T + A)$ is positive definite.

Proof of Claim 3:

If the matrix A is symmetric then the stated condition is exactly that of A being positive definite, which equivalently states that all eigenvalues of A are positive.

If A is not symmetric then take $A^* = \frac{1}{2}(A + A^T)$ which is symmetric.

Also, we clearly have

$$\mathbf{x}^T A\mathbf{x} \equiv (\mathbf{x}^T A\mathbf{x})^T = \mathbf{x}^T A^T\mathbf{x},$$

so that

$$\mathbf{x}^T A^*\mathbf{x} = \mathbf{x}^T A\mathbf{x}.$$

So from above we require A^* be positive definite.

2.4.3 Case: Diagonal Matrix A

We now move specifically to the 2-dimensional case, as it is the one modeled in Section 2.1.2. We also consider

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \alpha\sigma_1\sigma_2 \\ \alpha\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

with $\sigma_1, \sigma_2 > 0$ and $|\alpha| \leq 1$ (so this is a covariance matrix).

Our initial simplification of the process is to take A to be a diagonal matrix, after dealing with this case, we will relax the assumptions to more general cases. The physical interpretation of A being diagonal is that the two processes are attracted independently to 0, but deviations from the deterministic path there are correlated. Recall that we require the eigenvalues of A to be positive if the

process is to be ergodic, that is the diagonal entries of A are both positive. So we assume

$$A = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix},$$

with $\lambda_1, \lambda_2 > 0$. From (2.2) it is clear that

$$e^{At} = \begin{pmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{pmatrix}.$$

From (2.9), considering $\int_0^t e^{Ar}\Sigma d\mathbf{W}_r$, as the integrand is non-random, this is Gaussian with mean $\mathbf{0}$ and covariance given by, as both A and Σ are symmetric and e^{At} diagonal,

$$\int_0^t e^{Ar}\Sigma\Sigma^T e^{A^T r} dr = \int_0^t e^{Ar}\Sigma^2 e^{Ar} dr.$$

Letting $\sigma_1 = \sigma_2 = 1$, this reduces to

$$\begin{aligned} &\int_0^t \begin{pmatrix} e^{\lambda_1 r} & 0 \\ 0 & e^{\lambda_2 r} \end{pmatrix} \begin{pmatrix} 1+\alpha^2 & 2\alpha \\ 2\alpha & 1+\alpha^2 \end{pmatrix} \begin{pmatrix} e^{\lambda_1 r} & 0 \\ 0 & e^{\lambda_2 r} \end{pmatrix} dr \\ &= \begin{pmatrix} \frac{1+\alpha^2}{2\lambda_1}\left(e^{2\lambda_1 t}-1\right) & \frac{2\alpha}{\lambda_1+\lambda_2}\left(e^{(\lambda_1+\lambda_2)t}-1\right) \\ \frac{2\alpha}{\lambda_1+\lambda_2}\left(e^{(\lambda_1+\lambda_2)t}-1\right) & \frac{1+\alpha^2}{2\lambda_2}\left(e^{2\lambda_2 t}-1\right) \end{pmatrix}. \end{aligned}$$

The solution to $\mathbf{X}_t$ when $\mathbf{X}_0 = \mathbf{x}$ has the bivariate normal distribution given by

$$\mathbf{X}_t \sim \mathcal{N}_2\left(e^{-At}\mathbf{x}, e^{-At}\left(\int_0^t e^{Ar}\Sigma^2 e^{Ar} dr\right) e^{-At}\right).$$

In the case where $\sigma_1 = \sigma_2 = 1$ this reduces to

$$\mathcal{N}_2\left(\begin{pmatrix} x_1 e^{-\lambda_1 t} \\ x_2 e^{-\lambda_2 t} \end{pmatrix}, \begin{pmatrix} \frac{1+\alpha^2}{2\lambda_1}\left(1-e^{-2\lambda_1 t}\right) & \frac{2\alpha}{\lambda_1+\lambda_2}\left(1-e^{-(\lambda_1+\lambda_2)t}\right) \\ \frac{2\alpha}{\lambda_1+\lambda_2}\left(1-e^{-(\lambda_1+\lambda_2)t}\right) & \frac{1+\alpha^2}{2\lambda_2}\left(1-e^{-2\lambda_2 t}\right) \end{pmatrix}\right).$$

So that $\{\mathbf{X}_t\}$ has stationary distribution

$$\mathbf{X}_t \sim \mathcal{N}_2\left(\mathbf{0}, \begin{pmatrix} \frac{1+\alpha^2}{2\lambda_1} & \frac{2\alpha}{\lambda_1+\lambda_2} \\ \frac{2\alpha}{\lambda_1+\lambda_2} & \frac{1+\alpha^2}{2\lambda_2} \end{pmatrix}\right).$$

2.4.4 Case: Symmetric Matrix A

Relaxing the diagonal assumption and letting A be symmetric, we can view $A = PDP^{-1}$ where P is an orthonormal matrix, that is its columns consist of orthonormal vectors so that $P^T = P^{-1}$, and D is diagonal. In this case, $e^{At} = Pe^{Dt}P^{-1}$ and we continue as above to obtain

$$\int_0^t e^{Ar}\Sigma\Sigma^T e^{A^T r}dr = \int_0^t Pe^{Dr}P^T\Sigma^2 Pe^{Dr}P^T dr,$$

and when $\mathbf{X}_0 = \mathbf{x}$,

$$\mathbf{X}_t \sim \mathcal{N}_2\left(Pe^{-Dt}P^T\mathbf{x}, Pe^{-Dt}\left(\int_0^t e^{Dr}P^T\Sigma^2 Pe^{Dr}dr\right)e^{-Dt}P^T\right).$$

In terms of the physical nature of the process, this means that the extent to which the first component relates to the second is the same as the extent that second relates to the first.

2.4.5 Case: General Matrix A

For a general A, $e^{A^T t}$ need not equal e^{At}. However the process is still Gaussian and so has distribution given by

$$\mathbf{X}_t \sim \mathcal{N}_2\left(e^{-At}\mathbf{x}, e^{-At}\left(\int_0^t e^{Ar}\Sigma^2 e^{A^T r}dr\right)e^{-A^T t}\right)$$

but little else can be said.

2.5 Verification of Model Validity

From the analysis of the OU processes we are now in a position to show that the model properties in (1.3) are satisfied.

The univariate Ornstein-Uhlenbeck process is stationary when the initial distribution is constant and ergodic for $\rho > 0$ and so when this holds, the residuals for the first model, given by $\tilde{Y}_t$ have all the required properties.
Similarly for the multivariate process, when $\frac{1}{2}(A^T + A)$ has only positive eigenvalues, the process is ergodic and so will return to the line of interest sufficiently often, recall from model (2.3) that we are interested in when the OU process hits a line passing through the origin.

In this way our models are sensible for the problem at hand.

Chapter 3

Hitting Time Distributions

3.1 Preamble

Recall from Section 2.1 that we are going to consider two different models for movements of stocks X_t and Y_t. Our trading rule is to trade in the appropriate direction when the prices are sufficiently different from the historical price ratio and reverse the trade when equilibrium is re-established.

3.1.1 One Process

Recall from model (2.1) in Section 2.1 that in describing the price movements of two stocks X and Y we had factors F_t and a univariate OU process $\tilde{Y}_t$. X_t is fixed relative to F_t and Y_t fluctuates around it.

We set $t = 0$ when the trade begins, that is when $\log(Y_0) - \log(kX_0) = -\log(b)$ for some $b > 0$. The sign of $-\log(b)$ indicates which stock to short and which to long. We are interested in $\tau = \inf\{t \geq 0 : \log(Y_\tau) - \log(kX_\tau) = 0\}$.
Notice that $\log(Y_t) - \log(kX_t) = \tilde{Y}_t$. Thus the holding time of our trade (the time between opening and closing the trade) is:

$$\tau = \inf\{t \geq 0 : \tilde{Y}_t = 0\}, \qquad \tilde{Y}_0 = -\log(b).$$

As $\tilde{Y}_t$ is OU, this leads naturally to considering the hitting time of a level c by the OU process particularly $c = 0$.

3.1.2 Two Processes

Similarly recall from model (2.3) in Section 2.1 that we have our two stocks, X and Y with a common factor component, F_t. Each stock has an associated OU process which together form a bivariate OU process.

We initiate a trade at $t = 0$ where $\log(Y_0) - \log(kX_0) = -\log(b)$. The direction of the trade is determined by the sign of $-\log(b)$. We are concerned with the holding time of the trade, $\tau = \inf\{t \geq 0 : \log(Y_\tau) - \log(kX_\tau) = 0\}$.

Under this model, $\log(Y_t) - \log(kX_t) = \tilde{Y}_t - \tilde{X}_t$, so our holding time becomes

$$\tau = \inf\{t \geq 0 : \tilde{Y}_t - \tilde{X}_t = 0\}, \qquad \tilde{Y}_0 = \tilde{X}_0 - \log(b).$$

As $(\tilde{X}_t, \tilde{Y}_t)$ is bivariate OU process, this leads to analysing the hitting times of bivariate OU processes to linear boundaries.

3.2 Reflection Principle

The exposition in this section is based on the material presented in Chapter 13 of [7].

Theorem 5. [7, Theorem 13.11]: (Strong Markov property) For any Brownian motion $\{\mathbf{W_t}\}$ in $\mathbb{R}^d$ and finite (a.s.) stopping time τ, $\mathbf{W}'_t = \mathbf{W}_{t+\tau} - \mathbf{W}_\tau$,$t \geq 0$, is a Brownian motion independent of $\mathcal{F}_\tau$

Proof. Take a sequence $\tau_n \to \tau$ so that each τ_n takes only countably many values and $\tau_n \geq \tau + 2^{-n}$. Then $\mathcal{F}_\tau \subset \bigcap_n \mathcal{F}_{\tau_n}$.

Also, each $\mathbf{W}^n_t = \mathbf{W}_{\tau_n+t} - \mathbf{W}_{\tau_n}$ is a Brownian motion independent of $\mathcal{F}_\tau$ by the independent increments of Brownian motion, and $\mathbf{W}^n_t \to \mathbf{W}'_t$ (a.s) through continuity. Independence is guaranteed through dominated convergence and extension. □

Lemma 2. [7, Lemma 13.14]: Take a univariate Brownian motion W_t with an associated stopping time τ Then W_t and

$$\tilde{W}_t = W_{t\wedge\tau} - (W_t - W_{t\wedge\tau})$$

have the same distribution.

Proof. By the strong Markov property, $W'_t = W_{t+\tau} - W_\tau$ is a Brownian motion independent of τ and $W_{t\wedge\tau}$. Since W' is symmetric about 0,

$$(\tau, W_{t\wedge\tau}, W'_t) \stackrel{d}{=} (\tau, W_{t\wedge\tau}, -W'_t).$$

We notice that

$$W_t = W_{t\wedge\tau} + W'_{(t-\tau)^+}$$

and

$$\tilde{W}_t = W_{t\wedge\tau} - W'_{(t-\tau)^+}$$

so they have the same distribution. □

In particular:
Suppose $\{W_t\}$ is standard brownian motion, $W_0 = 0$ and $\tau_c = \inf\{t : W_t = c\}$, the first hitting time of level $c > 0$. Then

$$\begin{aligned}\mathbb{P}(W_t \geq c) &= \mathbb{P}(\tau_c \leq t; W_t - W_{\tau_c} \geq 0) \\ &= \mathbb{P}(\tau_c \leq t)\mathbb{P}(W_t - W_{\tau_c} \geq 0 | \tau_c \leq t),\end{aligned}$$

but $W_t - W_{\tau_c}$ is independent of its history (up to τ_c). This implies $(W_t - W_{\tau_c} | \tau_c \leq t)$ is just $W_t - W_{\tau_c}$ in distribution. This means that $\mathbb{P}(W_t - W_{\tau_c} \geq 0 | \tau_c \leq t) = \frac{1}{2}$. Thus

$$\mathbb{P}(\tau_c \leq t) = 2\mathbb{P}(W_t \geq c)$$

$$\begin{aligned}\mathbb{P}(\tau_c \leq t) &= \sqrt{\frac{2}{\pi t}} \int_c^\infty e^{-\frac{x^2}{2t}} dx \\ &= \sqrt{\frac{2}{\pi}} \int_{c/\sqrt{t}}^\infty e^{-\frac{y^2}{2}} dy.\end{aligned}$$

So τ_c has density given by

$$f_{\tau_c}(t) = \frac{c}{\sqrt{2\pi}} t^{-3/2} \exp\left\{-\frac{c^2}{2t}\right\}.$$

3.3 Time Change

Theorem 6. : (Dambis-Dubins-Schwarz)[4]
For X_t a continuous stochastic process with almost all paths nowhere constant, there exists some function Π such that $\Pi \circ X_t$ is a standard Drownian motion.

For a function $\pi(t)$, define

$$\Pi \circ X_t = X_{\pi(t)}$$

3.3.1 Time Changed Hitting Times

Univariate Process

An alternate way of dealing with the hitting problem is to use the above theorem to transform the OU process into a Weiner process. Recall for the univariate process that when $X_0 = x$, X_t has an explicit solution

$$X_t = e^{-\rho t}\left[x + \int_0^t \sigma e^{\rho r} dW_r\right].$$

Let us examine $\int_0^t \sigma e^{\rho r} dW_r$.

As the integrand is nonrandom this expression is Gaussian, has independent increments and has variance given by

$$\int_0^t (\sigma e^{\rho r})^2 dr = \frac{\sigma^2}{2\rho}(1 - e^{-2\rho t})$$

So if $s = \frac{\sigma^2}{2\rho}(1 - e^{-2\rho t})$ then $\int_0^t \sigma e^{\rho r} dW_r = \tilde{W}_s$ where $\tilde{W}_s$ is a standard Brownian motion.

Setting

$$s(t) = \frac{\sigma^2}{2\rho}(1 - e^{-2\rho t}),$$

we have

$$X_t = e^{-\rho t}[x + \tilde{W}_{s(t)}], \qquad \tilde{W}_{s(0)} = 0$$

Thus the case of X_t hitting level c (from $X_0 = x$) is the same as $\tilde{W}_{s(t)}$ hitting $ce^{-\rho t} - x$ starting from 0. Inverting $s(t)$ we have transformed the problem so

that we require $\tilde{W}_s, (\tilde{W}_0 = 0)$ to hit the boundary

$$c\sqrt{1 + \frac{2\rho s}{\sigma^2}} - x. \tag{3.1}$$

From this we infer that the density for X_t hitting level c from x is a transformation of $\tilde{W}_s$ hitting boundary in (3.1).

Note that if $\rho > 0$ this boundary is square root. It is well known that $\tilde{W}_s$ hits such boundaries with finite expectation, that is the process is positively recurrent. Sato, [15], gives tail behavior of the hitting time distribution.

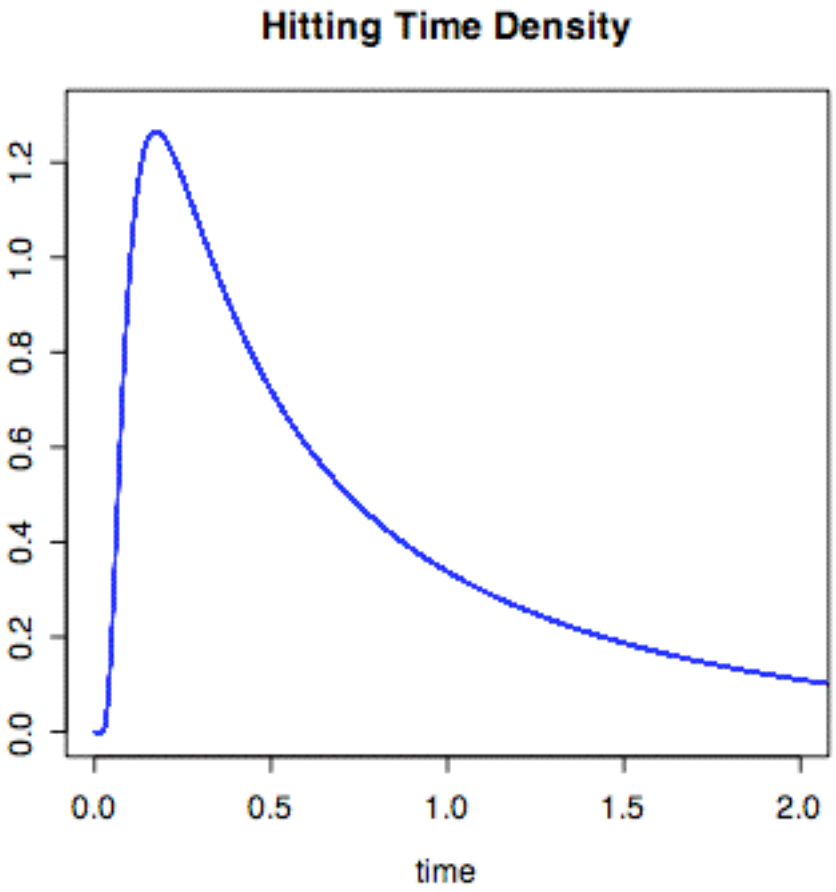
In the case that $c = 0$ we can find a closed form for the hitting density. The approach is outlined in [1]. When $c = 0$ our problem is reduced to a Weiner process hitting a level $-x$ which we know, from Section 3.2, the density to be:

$$f_{\tau_{-x}}(s) = \frac{|x|}{\sqrt{2\pi}} s^{-3/2} \exp\left\{-\frac{x^2}{2s}\right\}, \quad s > 0 \tag{3.2}$$

which means, transforming, we reach the well known result

$$f_\tau(t) = \frac{|x|}{\sqrt{2\pi}\sigma} \left(\frac{\rho}{\sinh(\rho t)}\right)^{3/2} \exp\left\{-\frac{\rho x^2 e^{-\rho t}}{2\sigma^2 \sinh(\rho t)} + \frac{\rho t}{2}\right\}, \quad t > 0.$$

A plot when $\rho = 1, \sigma^2 = 2$ is included below.



An alternate method for arriving at this density using 3 dimensional Bessel Bridges is given in [6]

Multivariate Process

We can use a similar procedure on the first component of a multivariate OU process so that after the transformation, it will become standard Brownian motion. However, this messes up the other components such that this process is of no help in solving the boundary hitting problem. The only case where it would help is where we have two independent OU processes, and in that case we can analyse them as separate univariate processes.

3.4 Differential Equation Approach

3.4.1 Univariate Process

Let $\tau = \inf\{t \geq 0 : X_t = c\}$. For the univariate process, $\mu(x,t) = -\rho x$ and $\sigma(x,t) = \sigma$.

From Section 1.4.4, setting $\psi(X_t, t) = e^{-\lambda t}$ it follows that $U(x) = \mathbb{E}\left[e^{-\lambda\tau}|X_0 = x\right]$ satisfies the differential equation

$$\rho x U'(x) - \frac{1}{2}\sigma^2 U''(x) = -\lambda U(x). \tag{3.3}$$

This differential equation has two linearly independent solutions given by $U_1(x) = D_{-\lambda/\rho}(x\sqrt{2\rho}/\sigma)$ and $U_2(x) = D_{-\lambda/\rho}(-x\sqrt{2\rho}/\sigma)$ where $D_\nu(x)$ is the parabolic cylinder function (Weber function).

The Laplace transform of the density of the first hitting to boundary boundary c from x was found in [3] and [2]. It is given by

$$\mathcal{L}_{x,c}(\lambda) = \begin{cases} \frac{D_{-\lambda/\rho}(-x\sqrt{2\rho}/\sigma)}{D_{-\lambda/\rho}(-c\sqrt{2\rho}/\sigma)} \exp\left(\frac{\rho(x^2-c^2)}{2\sigma^2}\right) & \text{if } x < c, \\ \frac{D_{-\lambda/\rho}(x\sqrt{2\rho}/\sigma)}{D_{-\lambda/\rho}(c\sqrt{2\rho}/\sigma)} \exp\left(\frac{\rho(x^2-c^2)}{2\sigma^2}\right) & \text{if } x > c. \end{cases}$$

The following derivation was given in [3].
As X_t is Markovian, letting $p(x|y,t)$ be the transition density from x to y in time t and $f_c(x|t)$ the density of τ and letting $\hat{p}$ and $\hat{f}$ being the Laplace transforms of these. By the famous Chapman-Kolmogorov equation

$$p(x|y,t) = \int_0^t f_c(x|r)p(c|y,t-r)dr.$$

When we take Laplace transforms of both sides, this becomes

$$\hat{p}(x|y,\lambda) = \hat{f}(x|\lambda)\hat{p}(c|y,\lambda).$$

From this we can see that $\hat{p}(x|y,\lambda) = u(x)v(y)$ for some u and v. So we see $\hat{f}_c(x|\lambda) = \frac{u(x)}{u(c)}$.

Then, as $\hat{f}_c(x|\lambda)$ solves (3.3), $u(x)$ must also be a solution to that differential equation. Since the solutions to the differential equation (3.3) are given, U_1, U_2 the Laplace transform is as given.

To extract the density from this, we need to invert this Laplace transform. It appears that no closed form solution exists in general, but numerical approximations can be obtained.

3.4.2 Multivariate Process

When dealing with a process in more than one dimension. We assume the process is governed by

$$d\mathbf{X}_t = \mu(\mathbf{X}_t, t)dt + \sigma(\mathbf{X}_t, t)d\mathbf{W}_t, \qquad \mathbf{X}_0 = \mathbf{x},$$

where $\sigma(\mathbf{X}_t, t)$ is a positive definite real $d \times d$ matrix.

A hyperplane in $d-$dimensional space is of $(d-1)-$dimensions. Any hyperplane is a rotation of another hyperplane with its first component fixed at some level. In view of this, we only consider hyperplanes constant in their first component when looking at the the hitting time problem.

The usual backward equation, Section 1.4.4 becomes

$$\frac{\partial}{\partial t}\psi(\mathbf{x}, t) = -\nabla_{\mathbf{x}}\psi(\mathbf{x}, t)^T\mu(\mathbf{x}, t) - \frac{1}{2}\mathrm{Tr}(\sigma(\mathbf{x}, t)H\sigma(\mathbf{x}, t)),$$

where H is the matrix of the second spatial derivatives of ψ and $\mathrm{Tr}(A)$ refers to the sum of the diagonal elements of A.

Now let $\tau = \inf\{t \geq 0 : \mathbf{X}_t^{(1)} = c\}$, $\mu(\mathbf{x}, t) = -A\mathbf{x}$, $\sigma(\mathbf{x}, t) = \Sigma$ and $u(\mathbf{X}_t, t) = e^{-\lambda t}$ so that we have an OU process. Then $U(\mathbf{x}) = \mathbb{E}\left[e^{-\lambda\tau}|\mathbf{X}_0 = \mathbf{x}\right]$ satisfies the equation

$$-\lambda U(\mathbf{x}) = \nabla U(\mathbf{x})^T A\mathbf{x} - \frac{1}{2}\mathrm{Tr}(\Sigma H\Sigma)$$

or equivalently, taking $A^{(i)}$ the ith column of A and $\Sigma^{(i,j)}$ the (i, j) entry of Σ,

$$-\lambda U(\mathbf{x}) - \sum_{i=1}^{d} A^{(i)}\mathbf{x}\, U_{x_i}(\mathbf{x}) + \frac{1}{2}\sum_{i,j=1}^{d}\Sigma^{(i,j)}U_{x_ix_j}(\mathbf{x}) = 0,$$

with boundary condition

$$U(c, x_2, \ldots, x_d) = 1.$$

This is an elliptic partial differential equation by positive definiteness of Σ. In non-trivial cases, a numerical approximation to the solution can be obtained.

3.4.3 Hitting Time Moments

All of the moments of the first hitting time can be computed from its Laplace transform by differentiaing the latter at zero:

$$\mathbb{E}[\tau^n|X_0 = x] = \frac{d^n}{d\lambda^n}(-1)^n \mathcal{L}_{x,c}(\lambda)\Big|_{\lambda=0}.$$

Following [14], we can rewrite $\mathcal{L}(\lambda)$ as

$$\mathcal{L}(\lambda) = \frac{\phi(\frac{\lambda}{\rho}, x\frac{\sqrt{2\rho}}{\sigma})}{\phi(\frac{\lambda}{\rho}, c\frac{\sqrt{2\rho}}{\sigma})},$$

where

$$\begin{aligned}\phi(a,b) &= \sum_{n=0}^{\infty} \frac{(\sqrt{2}b)^n}{n!} \frac{\Gamma(\frac{n+a}{2})}{\Gamma(\frac{a}{2})} \\ &= \sum_{n=0}^{\infty} \frac{(\sqrt{2}b)^n}{n!} \gamma_n(a), \end{aligned} \tag{3.4}$$

with $\gamma_n(a) = \frac{\Gamma((n+a)/2)}{\Gamma(a/2)}$.
Note that here we let $x < c$. Due to the symmetry of the process, x hitting c from above is the same as $-x$ hitting $-c$ from below. In this way, all cases are considered in looking at $x < c$.

Then, setting $\phi_k(b) = \frac{d^k}{d\lambda^k}\phi(\lambda, b)\big|_{\lambda=0}$, it follows that

$$\sum_{k=0}^{n} \binom{n}{k} (-1)^k \mathbb{E}[\tau^k|x] \frac{1}{\rho^{n-k}} \phi_{n-k}\left(c\frac{\sqrt{2\rho}}{\sigma}\right) = \frac{1}{\rho^n}\phi_n(x),$$

which in turn leads to

$$\mathbb{E}[\tau^n|x] = (-1)^n \left\{ \frac{1}{\phi^n}\phi_n\left(x\frac{\sqrt{2\rho}}{\sigma}\right) - \sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k} \mathbb{E}[\tau^{n-k}|x] \frac{1}{\rho^k}\phi_k\left(c\frac{\sqrt{2\rho}}{\sigma}\right) \right\}.$$

This a system of linear equations which can be solved in a recursive manner.

In the case where $\rho = 1$, $\sigma^2 = 2$ it reduces to

$$\mathbb{E}[\tau^n|x] = (-1)^n \left\{ \phi_n(x) - \sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k} \mathbb{E}[\tau^{n-k}|x]\phi_k(c) \right\}$$

Ricciardi and Sato [14] showed that the solution takes the form

$$\mathbb{E}[\tau^n|x] = \det\begin{pmatrix} \phi_1(c) & \phi_0(c) & 0 & 0 & \dots & 0 \\ \phi_2(c) & 2\phi_1(c) & \phi_0(c) & 0 & \dots & 0 \\ \phi_3(c) & 3\phi_2(c) & 3\phi_1(c) & \phi_0(c) & \dots & 0 \\ & & \dots & & & \\ \phi_n(x) & n\phi_{n-1}(c) & \binom{n}{2}\phi_{n-2}(c) & \binom{n}{3}\phi_{n-3}(c) & \dots & n\phi_1(c) \end{pmatrix}$$
$$- \det\begin{pmatrix} \phi_1(x) & \phi_0(c) & 0 & 0 & \dots & 0 \\ \phi_2(x) & 2\phi_1(c) & \phi_0(c) & 0 & \dots & 0 \\ \phi_3(x) & 3\phi_2(c) & 3\phi_1(c) & \phi_0(c) & \dots & 0 \\ & & \dots & & & \\ \phi_n(x) & n\phi_{n-1}(c) & \binom{n}{2}\phi_{n-2}(c) & \binom{n}{3}\phi_{n-3}(c) & \dots & n\phi_1(c) \end{pmatrix},$$

with $\phi_k(b) = \phi^{(k)}(0, b)$

It can also be shown, as in [14], that

$$\phi_k(b) = \frac{k}{2^k}\sum_{n=1}^{\infty}\frac{(\sqrt{2}b)^n}{n!}\Gamma\left(\frac{n}{2}\right)\alpha_n^{(k)},$$

with $\alpha_n^{(1)} = 1$ and $\alpha_n^{(k+2)}(k = 0, 1, 2, \dots)$ given by the following representation

$$\alpha_n^{(k+2)} = \det\begin{pmatrix} \psi_0 & -1 & 0 & 0 & \dots & 0 \\ \psi_1 & \psi_0 & -1 & 0 & \dots & 0 \\ \psi_2 & 2\psi_1 & \psi_0 & -1 & \dots & 0 \\ & & \dots & & & \\ \psi_k & k\psi_{k-1} & \binom{k}{2}\psi_{k-2} & \binom{k}{3}\psi_{k-3} & \dots & \psi_0 \end{pmatrix}, \tag{3.5}$$

where $\psi(z)$ is the digamma function, $\psi^{(k)}(z)$ the kth derivative of ψ evaluated at z and $\psi_k = \psi^{(k)}(\frac{n}{2}) - \psi^{(k)}(1)$.

$$\psi(z) = \frac{1}{\Gamma(z)}\frac{d}{dz}\Gamma(z)$$

To prove this, Given the representation in (3.4), we only need to show that

$$\frac{d^k}{da^k}\gamma_n(a)\big|_{a=0} = \frac{k}{2^k}\Gamma\left(\frac{n}{2}\right)\alpha_n^{(k)}.$$

If this is true, the power series coincide and equality of the series is assured.

Rewrite $\gamma_n(a)$ as $\frac{a}{2}\chi_n(a)$, where

$$\chi_n(a) = \frac{\Gamma(\frac{n+a}{2})}{\Gamma(1+\frac{a}{2})}.$$

From analyticity of $\chi_n(a)$ that all derivatives of χ_n exist in the vicinity of the origin, and so we can write

$$\gamma_n^{(k)}(a) = \frac{a}{2}\chi_n^{(k)}(a) + \frac{k}{2}\chi_n^{(k-1)}(a),$$

which implies that

$$\gamma_n^{(k)}(0) = \frac{k}{2}\chi_n^{(k-1)}(0).$$

Further, we have

$$\begin{aligned}
\frac{d}{da}\chi_n(a) &= \frac{\Gamma(1+\frac{a}{2})\frac{d}{da}\Gamma(\frac{n+a}{2}) - \Gamma(\frac{n+a}{2})\frac{d}{da}\Gamma(1+\frac{a}{2})}{2\Gamma(1+\frac{a}{2})^2} \\
&= \frac{\Gamma(\frac{n+a}{2})}{\Gamma(1+\frac{a}{2})}\frac{1}{2}\left[\frac{\frac{d}{da}\Gamma(\frac{n+a}{2})}{\Gamma(\frac{n+a}{2})} - \frac{\frac{d}{da}\Gamma(1+\frac{a}{2})}{\Gamma(1+\frac{a}{2})}\right] \\
&= \chi_n(a)\xi_n(a),
\end{aligned}$$

with

$$\xi_n(a) = \frac{1}{2}\left[\psi\left(\frac{n+a}{2}\right) - \psi\left(1+\frac{a}{2}\right)\right].$$

Differentiating this relation gives

$$\chi_n^{(k)}(0) = \sum_{i=0}^{k-1}\binom{k}{i}\chi_n^{(i)}(0)\xi_n^{(k-i)}(0),$$

from which it is clear that

$$\gamma_n^{(k)}(0) = \frac{k}{k-1}\sum_{i=0}^{k-1}\binom{k}{i}\gamma_n^{(i)}(0)\xi_n^{(k-i)}(0).$$

Then, if we let

$$\frac{d^k}{da^k}\gamma_n(0) = \frac{k}{2^k}\Gamma\left(\frac{n}{2}\right)\alpha_n^{(k)},$$

we can easily check that $\alpha_n(a)$ is given by the expression (3.5).

In particular,

$$\begin{aligned}
\mathbb{E}[\tau|x] &= \frac{1}{\rho}\left(\phi_1\left(c\frac{\sqrt{2\rho}}{\sigma}\right) - \phi_1\left(x\frac{\sqrt{2\rho}}{\sigma}\right)\right) \\
&= \frac{1}{2\rho}\sum_{n=1}^{\infty}\frac{2^n(\sqrt{\rho})^n(c^n - x^n)}{\sigma^n n!}\Gamma\left(\frac{n}{2}\right).
\end{aligned}$$

Or, in the standard case ($\rho = 1, \sigma^2 = 2$):

$$\mathbb{E}[\tau|x] = \frac{1}{2}\sum_{n=1}^{\infty} \frac{(\sqrt{2})^n(c^n - x^n)}{n!}\Gamma\left(\frac{n}{2}\right).$$

3.5 A Mixture Distribution Approximation

We have a differential equation whose solution is the Laplace transform of a density of interest. From this Laplace transform we can extract the moments of the distribution. We can use the knowledge of the first few moments of the distribution of τ to construct an approximation to it.

Definition 8. A *mixture* of random variables $\{Y_j\}_{j=1}^k$ with densities $\{f_j(y;\theta_j)\}_{j=1}^k$ is a random variable Y with density given by:

$$\begin{aligned} f(y;\theta) &= \sum_{j=1}^{k} c_j f_j(y;\theta_j), \\ c_j &\in [0,1] \quad \forall j \in \{1,\dots,k\}, \\ \sum_{j=1}^{k} c_j &= 1. \end{aligned}$$

Note that $f(y;\theta)$ is necessarily a probability density by the restrictions on c_j. Let the support of Y be S and the support of Y_j be S_j. Then trivially $S = \cup_j S_j$. Also, for all moments of Y we have

$$\begin{aligned} \mathbb{E}[Y^r] &= \int_S y^r f(y;\theta) dy \\ &= \int_S y^r \sum_{j=1}^{k} c_j f_j(y;\theta_j) dy \\ &= \sum_{j=1}^{k} \int_{S_j} y^r c_j f_j(y;\theta_j) dy \\ &= \sum_{j=1}^{k} c_j \mathbb{E}[Y_j^r]. \end{aligned}$$

Given a variable of interest τ, take an indexed family of random variables with parameter $\theta_j \in \mathbb{R}^{d_j}$, $\{Y_j\}_{j=1}^n$ with densities $f_j(y;\theta_j)$. Also define

$$\phi(j,r;\theta_j) = \mathbb{E}[Y_j^r]$$

Let Y be the mixture of Y_j which best approximates τ. A first and crude way of determining c_j and θ_j would be moment estimation. Given

$$m_r = \mathbb{E}[\tau^r]$$

we aim to find c_j and θ_j that satisfy

$$m_r = \sum_{j=1}^{k} c_j \phi(j, r; \theta_j), \quad r = 1, 2, \ldots, k, \tag{3.6}$$

for k being the minimum number of moments required to get a unique solution.

3.5.1 Distributions Used

In looking at the hitting times of the OU process we are dealing with a non-negative, univariate quantity so an appropriate family of distributions to use could be Gamma distributions, $\gamma(k_j, \theta_j)$, with scale parameters, θ_j, and shape parameters, k_j, $\theta_j, k_j \in \mathbb{R}$

For $Y \sim \gamma(k, \theta)$

$$\mathbb{E}[Y^r] = \theta^r \frac{\Gamma(k+r)}{\Gamma(k)}, \quad k = 1, 2, \ldots$$

A sensible method for dealing with a hitting time density would be to only consider gamma distributions which have zero density at $t = 0$ so that at least this property of the actual hitting time is preserved, so $k = 1$ would not be a good model.

In what follows, let $\gamma(y; k, \theta)$ be the density for the random variable with gamma distributions with scale parameter θ and shape parameter k

3.5.2 Comparison With Simulations

Below is some R code used to generate 10^4 first hitting times of a level by a univariate Ornstein-Uhlenbeck process. We considered the case

$$\rho = 1, \quad \sigma = \sqrt{2}, \quad x =, 1 \quad c = 0$$

We choose this case because it has a well known density, and so we can compare the mixture approximations to the actual density to determine their validity. Then, when the exact hitting density is unknown, we have an idea of what form the mixture would take for it to be a decent approximation.

```
> ######################################################
> ###          set estimation parameters           ###
> ######################################################
> h <- 0.005
> n<- 10000
> ######################################################
> ###      set the parameters of the process       ###
> ###     dX(t) = -rho X(t) dt + sigma dW(t)       ###
> ######################################################
> rho <- 1
> sigma <- sqrt(2)
> hit <- 0           ###   hitting level 0         ###
> xinit <- 1         ###   starting from X(0)=1    ###
> ######################################################
> ###   generate n realisations of the OU process  ###
> ######################################################
> kappa <- exp(-rho*h)
> zvar <- 0.5*sigma^2/rho
> times <- 0
> for(i in 1:n){
+   t <- 0
+   x<- xinit
+   while((hit-x)*(hit-xinit) >= 0){
+     t <- t+h
+     x<- kappa*x + sqrt(1-kappa^2)*rnorm(1,0,zvar)
+     }
+   times[i]<- t    ###   vector of hitting times  ###
+   }
```

This gave the following approximate values for the moments:

$$\mathbb{E}(\tau) \approx 0.978147$$
$$\mathbb{E}(\tau^2) \approx 1.943$$
$$\mathbb{E}(\tau^3) \approx 6.03113$$
$$\mathbb{E}(\tau^4) \approx 24.97828$$

If we compare these moments to those found using Section 3.4.3 we find that the mean is larger in the simulations than Section 3.4.3 would suggest. This indicates a bias in the discrete approximation used, it tends to give slightly higher hitting times compared to the true values.

From (3.2) we know that the density of this hitting time is equal to

$$f_\tau(t) = \frac{1}{2\sqrt{\pi}} \left(\frac{1}{\sinh(t)} \right)^{3/2} \exp\left\{ -\frac{e^{-t}}{4\sinh(t)} + \frac{t}{2} \right\}, \quad t \geq 0.$$

We now fit various mixtures to the data and plot the resulting approximate density against this known density, keeping in mind that the simulation has a bias so that the hitting times are slightly larger than they should be. In trying to fit moments as in (3.6), the equations were inconsistent in many cases and so no solution was obtained.

Because our simulation generates 10^4 observations, we can view it as a sample and fit the appropriate mixtures to this 'sample'. We use *maximum likelihood* to determine the parameter estimates. This takes the estimated c_j, θ_j and k_j to be those values under which the observed sample is most likely.

What follows is meant to give an idea of the validity of the various gamma approximations rather than to be statistically rigorous. Given an approximation $\hat{f}_\tau(t)$ to $f_\tau(t)$, we would like the quadratic error

$$\int_0^\infty \left(\hat{f}_\tau(t) - f_\tau(t) \right)^2 dt$$

to be as small as possible. However, we will instead look at graphs of the mixture densities and comment on the validity (or not) of an approximation through this inexact visual measure.

In what follows, maximum likelihood estimates have been obtained in varying frameworks. The α_i mentioned are θ_i^{-1} in our formulation.

1. Fitting a single gamma distribution with no parameter restrictions

```
> llh <- function(pars){
+  k<-pars[1]
+  alpha<-pars[2]
+  -sum(dgamma(times,k,alpha,log=TRUE))}
> optim(c(1,1),llh)
$par
k=1.273920
alpha=1.302496
```

2. Two gamma distributions with no restrictions on the parameters

```
> llh <- function(pars){
+  alpha1<-pars[1]
+  k1<-pars[2]
+  alpha2<-pars[3]
+  k2<-pars[4]
+  c<-pars[5]
+  -sum(log(c*dgamma(times,k1,alpha1)+(1-c)*dgamma(times,k2,alpha2)))}
> optim(c(1,1,2,2,0.5),llh)
$par
alpha1=1.3080598
k1=1.8247097
alpha2=8.0589896
k2=2.9502948
c=0.5973547
```

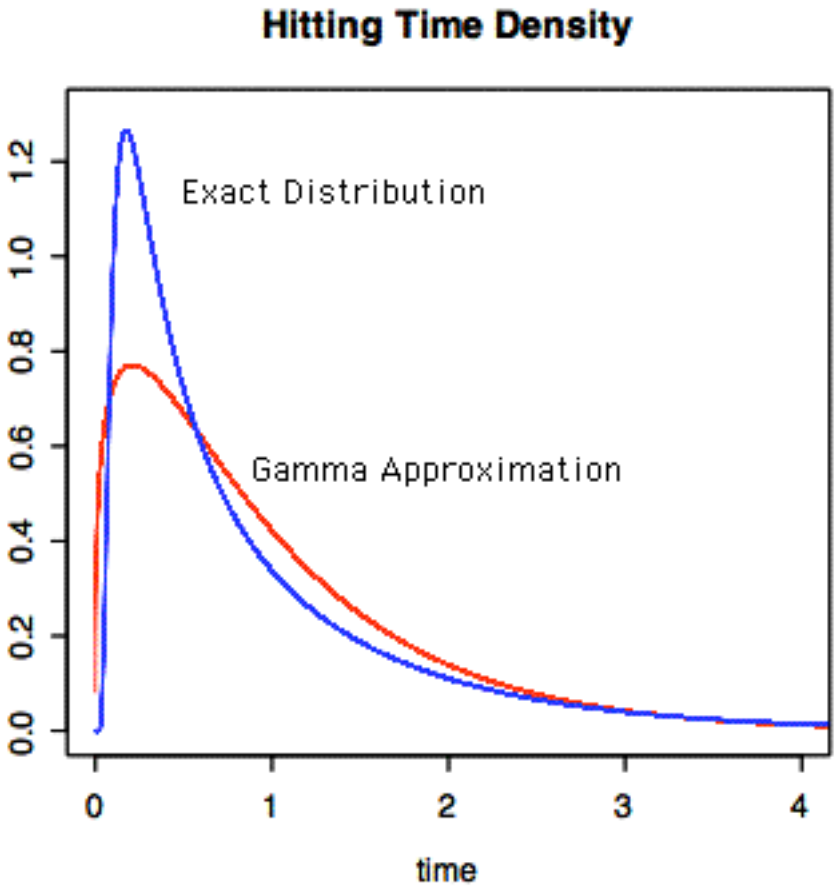
3. Three gamma distributions with no restrictions on the parameters

```
> llh <- function(pars){
+  alpha1<-pars[1]
+  k1<-pars[2]
+  alpha2<-pars[3]
```

```
+  k2<-pars[4]
+  alpha3<-pars[5]
+  k3<-pars[6]
+  c1<-pars[7]
+  c2<-pars[8]
+  -sum(log(c1*dgamma(times,k1,alpha1)+c2*dgamma(times,k2,alpha2)+(1-c1-
> pars<-optim(c(1,1,2,2,3,3,0.3,0.4),llh)
>
> pars
$par
alpha1=1.1053022
k1=2.5650427
alpha2=4.2744285
k2=2.8428879
alpha3=5.7225660
k3=3.9676221
c1=0.2371395
c2=2.1259410
```

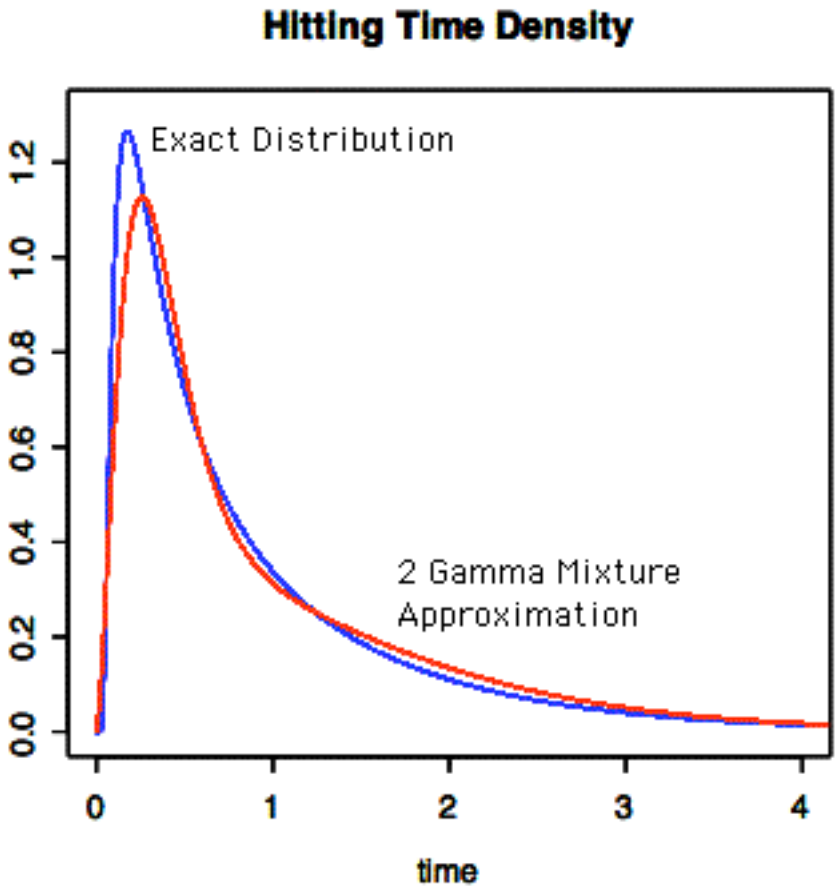
Plots of the hitting time density and the density from these gamma approximations are included below:

The fit of a single gamma distribution with no parameter restrictions is:



This is a very poor approximation, but is only really a starting point. One good point is that the right tail appears not too bad which indicates that gamma distributions are not an altogether terrible choice.

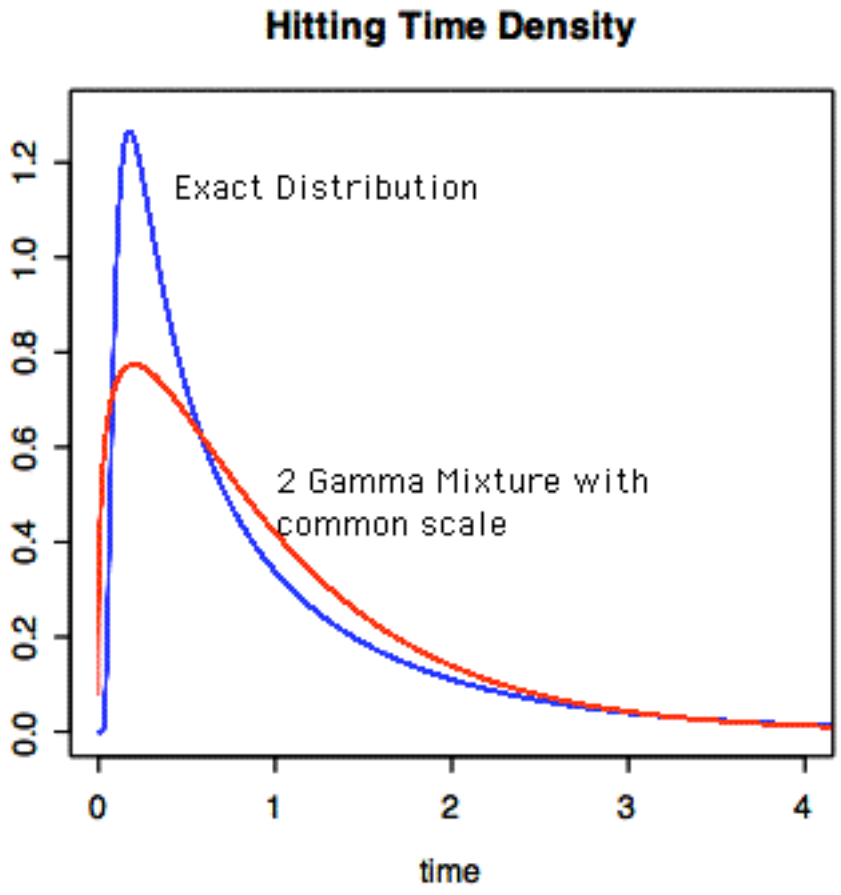
A mixture of two gamma distributions with no parameter restrictions gives a fit:



This is already a much better approximation. Considering the slight bias from the simulation means that it is even closer to the real distribution than it looks.

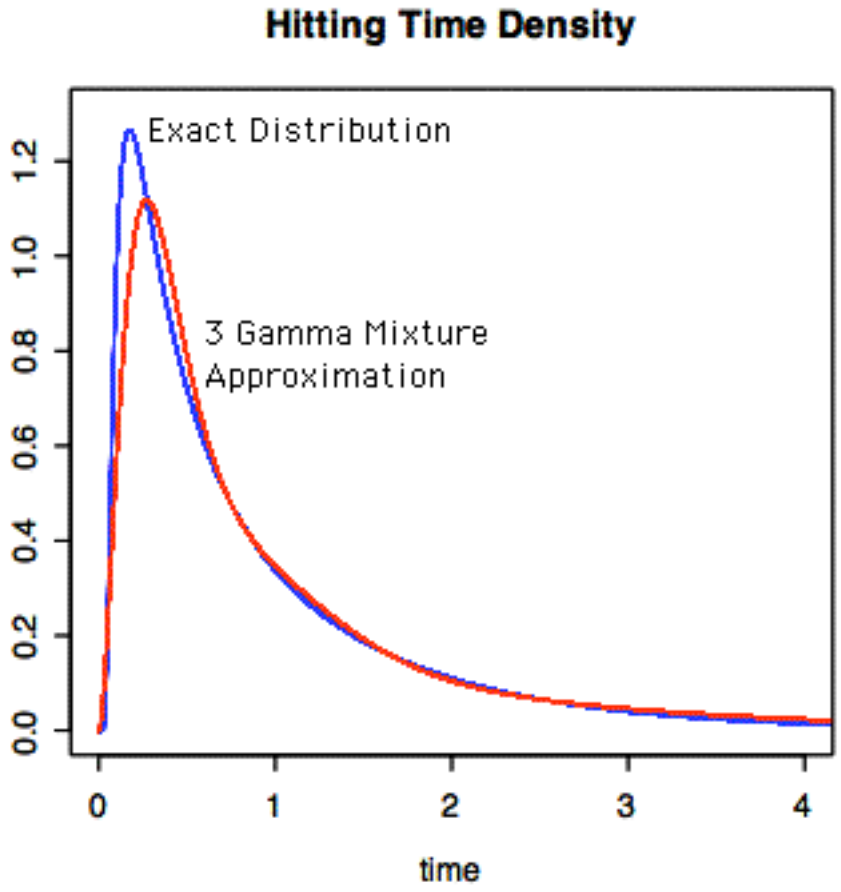
Another problem is that the optimisation is sensitive to the initial parameters used for the 'optim' function in R. However, in all variation in the parameters obtained, the graph of the density did not vary noticeably.

A mixture of two gamma distributions with the scale parameters constrained to be the same for both distributions gives a fit:



When we restrict that $\theta_1 = \theta_2$, to find the optimal mixture of gamma distributions with common scale (and rate), we get a poor approximation to the actual distribution. Thus, when we consider a two gamma mixture, we will consider two completely distinct gamma distributions.

A mixture of three gamma distributions with no parameter restrictions gives a fit:



This is also a good fit but does not appear to differ from the two distribution approximation greatly. Part of the problem may be that in the parameter estimates one of the weightings is negative and another is greater than 1. The c_i could be constrained so that $c_i \in [0, 1]$, but this can only have the effect of making the fit worse. Since the unconstrained three gamma distribution mixture is not greatly better than the two distribution, the constrained mixture will not be either.

Again we find that when different initial parameters (for 'optim') are used,

the parameter estimates vary quite a lot although the resulting graph does not change greatly.

Remarks

It would appear from this that a two gamma distribution mixture provides a decent approximation to the density in this known case and so this type of mixture may be appropriate in the situations where no closed form for the hitting density exists, (i.e. not hitting level 0) and for the multivariate hitting time.

Chapter 4

Concluding Remarks

The trades described in Sections 2.1.1 and 2.1.2 generate a positive net cashflow at time 0 and at their conclusion generate zero cashflow. This means that the only risk we are exposed to is that the stocks will never attain the required ratio. But we know from Chapter 2 that the stocks return to this ratio with probability 1. These trades are the ones we will be considering, since exit conditions and the ratio of stocks to trade were specified for them, so all that remains is to determine optimal entry conditions. That is, we need to determine b such that when the price ratio hits level b, we trade as described in Sections 2.1.1 and 2.1.2.

In this chapter we look at ways to develop an optimal trade. Optimality in this sense means that time taken per dollar earned would be as small as possible.

4.1 One Process Model

When we begin at a ratio $b > 1$, the returns are given by $Y_0(1-b) = kX_0(b^{-1}-1)$ and the expected hitting time of hitting level 0 from $\log(b)$ is given by

$$\frac{1}{\rho}\left(\phi_1(0) - \phi_1\left(\log(b)\frac{\sqrt{2\rho}}{\sigma}\right)\right),$$

where $\phi_k(z)$ are as defined in Section 3.4.3.

When the ratio b is large, to enter the trade, we are waiting for the OU process to reach some high level. This will happen very infrequently and so violates our requirement in Section 2.1 that trades would be in the short term so that the

time value of money could be ignored. Let B_U be the largest value for b such that the frequency for the ratio of prices to exceed b will be high enough from a practitioner's standpoint.

A ratio of b close to 1 would mean we enter a trade at a very slight deviation for the process. This would entail in a large amount of trades over a very short time, resulting in transaction costs becoming greater than neglibible. This violates the requirements of Section 2.1. Let B_L be the least value for b such that the frequency of trades results in a rate of transaction costs that are not too high from a practitioner's standpoint.

The problem reduces to finding

$$\begin{aligned} z^* &= \min_{B_L \leq b^* \leq B_U} z(b) \\ &= \min_{B_L \leq b^* \leq B_U} \left\{ \frac{\phi_1(0) - \phi_1\left(\log(b)\frac{\sqrt{2\rho}}{\sigma}\right)}{\rho Y_0(b-1)} \right\}, \end{aligned}$$

The problem now clearly amounts to finding b^* solving the equation for the critical point

$$b^*\sigma\sqrt{\rho}\left(\phi_1\left(\log(b^*)\frac{\sqrt{2\rho}}{\sigma}\right) - 1\right) = \sqrt{2}(b^*-1)\phi_2\left(\log(b^*)\frac{\sqrt{2\rho}}{\sigma}\right), \quad B_L \leq b^* \leq B_U$$

and taking the minimum of $z(B_L)$, $z(B_U)$ and $z(b^*)$.

This equation could be solves numerically.

4.2 Two Processes Model

For the model described in (2.3), while we have an expression for the cashflow generated by the trade, we can only approximate the expected holding time for the trade. Given the Laplace transform of the density of τ_b, the time taken for the the ratio $\frac{Y}{kX}$ to reach 1 from b, we can obtain the expected value and thus choose the optimal ratio b. This is a possible direction for further work.

4.3 Further Work

Some things that could be done from here are:

- Numerically solve the partial differential equations for the Laplace transform of multivariate OU process hitting time, and numerically invert this Laplace transform.
- Given some real data for stocks for which a pairs trade is feasible, estimate parameters for both the one process and two processes models. Then devise and run a formal test to see if the two processes model is significantly better than the one process model.
- Use a better estimation procedure for the mixture distribution than the method of moments. The EM algorithm is particularly useful in this sort of problem, see for example [9].
- Use a wider variety of distributions for the mixture approximation method.
- Devise and perform a formal test for determining which of the models from Section 3.5.2 is best for the density in question.

Bibliography

[1] L. Alili, P. Patie, and J.L. Pederson. Representations of the first hitting time density of an Ornstein-Uhlenbeck process. *Stochastic Models*, 21:967–980, 2005.

[2] I.F. Blake and W.C. Lindsey. Level crossing problems for random processes. *IEEE Transactions on Information Theory*, IT-19(3):295–315, May 1973.

[3] D.A. Darling and A.J.F. Siegert. The first passage problem for a continuous Markov process. 24:624–639, 1953.

[4] L.E. Dubins and G. Schwarz. On continuous martingales. *Proceedings of the National Academy of Sciences*, 53(5):913–916, May 1965.

[5] D.S. Ehrmann. *The Handbook of Pairs Trading.* John Wiley & Sons, Hoboken, New Jersey, 2006.

[6] A. Göing-Jaeschke and M. Yor. A clarification note about hitting times densities for Ornstein-Uhlenbeck processes. *Finance and Stochastics*, 7:413–415, 2003.

[7] O Kallenberg. *Foundations of Modern Probability.* Springer, New York, 2002.

[8] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus.* Springer-Verlag, New York, 1988.

[9] D. Karlis and E. Xekalaki. Choosing initial values for the EM algorithm for finite mixtures. *Computational Statistics & Data Analysis*, 41:577–590, 2003.

[10] F.C. Klebaner. *Introduction to Stochastic Calculus With Applications.* Imperial College Press, London, 2005.

[11] P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag, Berlin, 1992.

[12] P. Lancaster. *Theory of Matrices*. Academic Press, New York, 1969.

[13] Y. Lin and M. McCrae. The effects of $I(1)$ series on cointegration. *Journal of Applied Mathematics and Decision Sciences*, 6(4):229–240, 2002.

[14] L.M. Ricciardi and S. Sato. First-passage-time density and moments of the Ornstein-Uhlenbeck process. *Journal of Applied Probability*, 25(1):43–57, Mar 1988.

[15] S. Sato. Evaluation of the first-passage time probability to a square root boundary for the Wiener process. *Journal of Applied Probability*, 14(4):850–856, 1977.

[16] M. Vidyamurthy. *Pairs Trading, Quantitative Methods and Analysis*. John Wiley & Sons, Hoboken, New Jersey, 2004.