Derivative Securities, Fall 2010
Mathematics in Finance Program
Courant Institute of Mathematical Sciences, NYU
Jonathan Goodman
http://www.math.nyu.edu/faculty/goodman/teaching/DerivSec10/resources.html

Week 10

1 More on the Gaussian copula model

There is a simple general way to map a one dimensional random variable, X, to a one dimensional standard normal, Y. This is built on a transformation that takes either one to a standard uniform random variable, U. Standard uniform means that the probability density of U is $h(u) = 1$ if $0 \leq u \leq 1$, and $h)u) = 0$ otherwise. Random number generators produce (approximately) independent standard uniforms. Monte Carlo algorithms like *Box Muller*, *mapping*, and *rejection* create random variables with other distributions out of standard uniforms. The copula construction is based on mapping.

The cumulative distribution function (CDF), $F(x) = \Pr(X \leq x)$ maps any random variable, X to a standard uniform. If X is a random variable whose CDF is F, then

$$U = F(X) \tag{1}$$

has the standard uniform distribution. This goes both ways. If U is a standard uniform and X is found from U by solving (1) for X, then X has the CDF F.

The relationship between X and U is easiest to understand when the probability density $f(x)$ never vanishes. In that case $F'(x) = f(x) > 0$ for all x. This implies that for any U there is a unique X that satisfies (1). If the probability density $f(x)$ has no $\delta-$function components, then $F(x)$ is continuous and U is well defined for each X. Both of these conditions are violated in practical applications we are interested in. For example, the exponential random variable has probability density $f(x) = \lambda e^{-\lambda x}$ if $x \geq 0$ and $f(x) = 0$ if $x < 0$. However, it is easy to see that (1) defines a unique relationship between U and X in this case too. The Bernoulli random variable has $X = 0$ with probability p and $X = 1$ with probability $1 - p$. Its probability density is $f(x) = p\delta(x) + (1 - p)\delta(x - 1)$. In that case, the relationship (1) becomes $X = 0$ if $0 \leq U < p$, and $X = 1$ if $p \leq U \leq 1$. If we know U then X is determined. If we know X, then U is still to some extent random. The point is that it is easy to construct copulas in cases that do not have $0 < f(x) < \infty$ for all x. But the construction may be more complicated to describe in general.

The verification of the copula facts is so simple it can be confusing. Let us check first that if X has F as its CDF and U is found from X using (1), then U is uniformly distributed in the interval $[0, 1]$. Clearly, $U \in [0, 1]$, because $F(x)$ is a probability and therefore is in the interval $[0, 1]$ (I said this was easy). To show that U is uniformly distributed, we need to show that for any $u \in [0, 1]$, $\Pr(U < u) = u$. To do that, let x be the unique number with $F(x) = u$. Then

$U \leq u$ is equivalent to $X \leq x$. So $\Pr(U \leq u) = \Pr(X \leq x) = F(x) = u$, Q.E.D.[1] Suppose, on the other hand, that U is uniformly distributed in $[0,1]$ and that X is found from U by solving (1). It is as easy to see that $\Pr(X \leq x) = F(x)$. Indeed, if $u = F(x)$, then we know $\Pr(U < u) = u = F(x)$. Therefore, $\Pr(X \leq x) = \Pr(U \leq u) = u\ f(x)$, Q.E.D, again.

Using the two way nature of this mapping, we can transform any one dimensional random variable into any other. If F is the CDF of X and G is another CDF, then we can start with X, calculate U from (1), then calculate Y by solving $G(Y) = U$. This makes Y a function of X, and G is the CDF of Y. We can avoid writing U (but not calculating it) by writing these two steps as

$$X \iff Y \quad \text{if} \quad G(Y) = F(X)\ .$$

With this relation we can calculate X from Y or Y from X. A special case is when Y is a standard normal and $G(y) = N(y)$, then

$$X \iff Y \quad \text{if} \quad N(Y) = F(X)\ . \tag{2}$$

Let us make this concrete by combining it with the one factor correlation model from last week. In this model, there are idiosyncratic factors Z_i for each Y_i and one common factor, Z_0. The Z_k are i.i.d $\mathcal{N}(0,1)$ standard normals. The Y_i come from these *factor loadings* a_i in the interval $[-1,1]$ using

$$Y_i \;=\; a_i Z_0 \;+\; \sqrt{1-a_i^2}\, Z_i\ . \tag{3}$$

This has the consequence that $\text{var}(Y_i) = a_i^2 + (1-a_i^2) = 1$, and

$$\begin{aligned} \text{cov}(Y_i,Y_j) \;=\; E[Y_iY_j] \;&=\; E\left[\left(a_iZ_0 + \sqrt{1-a_i^2}Z_i\right)\left(a_jZ_0 \;+ \sqrt{1-a_j^2}Z_j\right)\right] \\ &=\; a_ia_jE\left[Z_0^2\right] \;=\; a_ia_j\ , \end{aligned}$$

if $i \neq j$. The covariance matrix of the Y_i in this one factor model is the sum of a diagonal matrix and a matrix of rank one. The diagonal matrix has $1 - a_i^2$ on the diagonals. The rank one[2] matrix is aa^t, so the total covariance matrix is

$$C \;=\; \text{cov}(Y) \;=\; E\left[YY^t\right] \;=\; \text{diag}(1-a_i^2) \;+\; aa^t\ .$$

Let us use this to create correlated default times in the simple exponential default model. In that model, T_i is the default time for bond i. Suppose we have

[1] Q.E.D. stands for the Latin phrase "quod (*that which*) erat (*was supposed to be*) demonstrandum (*proven*)". We use the phrase now to announce the end of a proof, even an easy one like this. European math and science books used to be written in Latin. Newton's *Philosophi Naturalis Principia Mathematica* was published in 1687. Gauss' *Disquisitiones Arithmeticae* was published in 1801.

[2] The *rank* of a matrix is the dimension of the vector space spanned by its columns, which is the same (a theorem of linear algebra) as the dimension of the vector space spanned by the rows. If a matrix, M, has rank one, then the columns are proportional to each other. This means that that the columns satisfy $m_2 = a_2m_1$, $m_3 = a_3m_1$. If $a = (1, a_2, \ldots, a_n$ is the row vector, this says that $M = m_1a$. A symmetric matrix of rank one must have the form aa^t, for some column vector a.

a pool of n bonds with the same rating and maturity. If λ is the common default intensity, then the CDF for each of the T_i is $F(t) = \Pr(T_i < t) = 1 - e^{-\lambda t}$. The solution to $F(T) = U$ is $T = -\frac{1}{\lambda}\ln((1-U)$. The general copula formula in this case boils down to

$$T_i = \frac{-1}{\lambda}\ln\left(1 - N(Y_i)\right) . \tag{4}$$

Note that $1 - U = 1 - N(Y)$ is larger than zero, so the log is defined, and is less than one, so the log is negative. Therefore, (4) produces a positive T.

There is the question of how to choose the factor loadings a_i. If the bonds really are identical, one would choose them all to be the same. Taking $a_i = \sqrt{\rho}$ makes ρ the common correlation coefficient between any pair Y_i, Y_j. With this, the complete correlated default model has just three parameters, λ, ρ, and the recovery rate, R.

There is a formula more general than (3) that allows you to construct correlated Gaussian random variables with any positive definite covariance matrix, C. One way to do it is to use the *Cholesky factorization*, $C = LL^t$. The Cholesky factorization theorem is that any positive definite symmetric matrix C may be factored as $C = LL^t$, where L is an *upper triangular* matrix. Upper triangular means that the entries of L below the diagonal are all zero: $L_{ij} = 0$ if $j < i$. Any good linear algebra software package can compute L from C in about as much work as it takes to solve a system of linear equations. For a 500×500 matrix, C, the Clolesky factorization takes less than a second on my several years old midrange laptop using LAPACK. Once you have L, you form a column vector, Z of independent standard normals. The covariance matrix of Z is $C_Z = E[ZZ^t] = I$. Then you compute $Y = LZ$. The covariance of Y then becomes

$$C_Y = E\left[YY^t\right] = E\left[LZZ^tL^t\right] = LE\left[ZZ^t\right]L^t = LIL^t = LL^t = C .$$

As an example, one might try to use sector factors as well as an overall factor to model bond defaults. For example, you could classify the bonds into sectors (housing, tech, energy, etc.) and create sector factors as well as an overall factor. Then each bond could have a sector loading as well as the idiosyncratic and overall market factors. It is not clear how one would calibrate such a model.

2 Market price of risk

We now leave credit risk and begin a long discussion of *short rate* models of the yield curve. In these models, the yield curve is explained using a single diffusion process for r_t, the short rate, or *overnight rate*. In these models the short time bond price is given by

$$B(t, t+dt) = e^{-r_t dt} . \tag{5}$$

At time $t = 0$ (today), the future short rate r_t is unknown. Only $B(0, dt) = e^{-r_0 dt}$ is known today. A primary goal of this approach to interest rate modeling is to predict the yield curve, deriving a formula for $B(0, t)$, from the stochastic differential equation for r_t.

The arguments here are different from those used to derive the Black Scholes model because they do not require r_t to be a tradable asset. The short rate *one factor* models assume that the yield curve at any given time t is a function of r_t alone. This means that in the part of the economy related to the yield curve, every tradable asset is driven by the same single source of noise that drives r_t. However, the various traded assets are not payouts on r_t, as they were in the Black Scholes derivative pricing model.

Following Hull, Chapter 27, suppose the traded assets have price histories f_t, g_t, h_t, etc. Given that prices are positive, we can write the dynamics as

$$df_t = \mu_{f,t} f_t \, dt + \sigma_{f,t} f_t \, dW_t \, , \tag{6}$$

$$dg_t = \mu_{g,t} g_t \, dt + \sigma_{g,t} g_t \, dW_t \, , \tag{7}$$

etc. Two comments about this. One is that these really are the definitions of $\mu_{f,t}$ and $\sigma_{f,t}$. We do not suppose that μ and σ are constant. We do not know how they change with time or how they differ from each other. But if f_t is a diffusion and is positive, we can define $\mu_{f,t}$ and $\sigma_{f,t}$ as

$$\mu_{f,t} = \frac{E\left[df_t \mid \mathcal{F}_t\right]}{f_t \, dt} \, ,$$

$$\sigma^2_{f,t} = \frac{\mathrm{var}\left[df_t \mid \mathcal{F}_t\right]}{f_t^2 \, dt} \, .$$

Economists have the notion of risk vs. return as a tradeoff – you get more expected return in exchange for accepting more risk. Of course, some investors are more risk averse than others. But the market should come to some balance where the law of supply and demand turns the risk preferences of the market participants into an overall market risk aversion. One manifestation of this is the quantity *market price of risk*, which is

$$\lambda_t = \frac{\mu_{f,t} - r_t}{\sigma_{f,t}} \, . \tag{8}$$

The notation here has λ depending on time, but not on f. This is the fundamental theorem of one factor markets: The market price of risk is the same for every traded asset in a one factor market.

The proof uses a hedging argument of the kind we used in the Black Scholes pricing theory. We make a locally risk free portfolio that is carefully constructed linear combination of traded assets, and argue that the return on risk free portfolios is the risk free rate. For now, everything takes place at the same time t, so I do not write the t subscript. Let f and g be two traded assets, and choose a portfolio

$$\Pi = \Delta_f f - \Delta_g g \, .$$

The weights $\Delta_f = g\sigma_g$ and $\Delta_g = f\sigma_f$ make the coefficient of dW in $d\Pi$, which is

$$\Delta_f f \sigma_f - \Delta_g g \sigma_g = g\sigma_g f \sigma_f - f\sigma_f g \sigma_g$$

equal to zero. Therefore, $r\Pi\,dt = E[d\Pi]$, which gives

$$r\,(\,g\sigma_g f \;-\; f\sigma_f g\,)\;dt \;=\; \Delta_f E\,[df] \;-\; \Delta_g E\,[dg] \;=\; g\sigma_g\mu_f f\,dt \;-\; f\sigma_f\mu_g g\,dt\;.$$

Removing the common factors $fg\,dt$ simplifies this to

$$r\sigma_g \;-\; r\sigma_f \;=\; \mu_f\sigma_g \;-\; \mu_g\sigma_f\;,$$

which may be manipulated to

$$(\mu_f - r)\,\sigma_g \;=\; (\mu_g - r)\,\sigma_f\;,$$

and finally to

$$\frac{\mu_f - r}{\sigma_f} \;=\; \frac{\mu_g - r}{\sigma_g}\;.$$

This proves the claim that the right side of (8) is independent of the asset f.

The Black Scholes theory led us to consider reweightings of the measure defining the process S_t that change the expected return from μ to r. In general, we saw that Girsanov's theorem allows reweightings that change the drift coefficient, $a(x)$, but not the noise coefficient $b(x)$ in a diffusion given by $dX = a(X)dt \,+\, b(X)dW$. In particular, we saw that the weight

$$L \;=\; e^{\int_0^T m_t dW_t \,-\, \frac{1}{2}\int_0^T m_t^2 dt}$$

Makes

$$E_L\,[dW_t \mid \mathcal{F}_t] \;=\; m_t dt$$

while keeping $E_L\left[dW_t^2\right] = dt$. In this weighted world, we can calculate

$$E_L\,[df_t \mid \mathcal{F}_t] \;=\; \mu_{f,t} f_t dt \,+\, \sigma_{f,t} f_t E_L\,[dW_t] \;=\; \mu_{f,t} f_t dt \,+\, \sigma_{f,t} f_t m_t dt\;. \qquad (9)$$

Of course, the short time variance is unchanged:

$$\mathrm{var}_L\,[df_t \mid \mathcal{F}_t] \;=\; \sigma_{f,t}^2 f_t^2\,dt\;. \qquad (10)$$

This means that in the reweighted world, the equations (6) and (7) are changed to

$$df_t \;=\; (\mu_{f,t} \,+\, \sigma_{f,t} m_t)\,f_t\,dt \,+\, \sigma_{f,t} f_t\,d\widetilde{W}_t\;, \qquad (11)$$

$$dg_t \;=\; (\mu_{g,t} \,+\, \sigma_{g,t} m_t)\,g_t\,dt \,+\, \sigma_{g,t} g_t\,d\widetilde{W}_t\;. \qquad (12)$$

The notation $\widetilde{dW}$ indicates that the Brownian motion in (11) and (12) is different from the one in (6) and (7). But you are not supposed to take this so seriously. The real definition of a stochastic process is (9) and (10). The stochastic differential equation (11) is just a convenient way to express (9) and (10).

Well, the equations (11) and (12), in that it is the same $\widetilde{dW}$ in both, also express the fact that

$$\mathrm{cov}(df_t, dg_t \mid \mathcal{F}_t) \;=\; E_L[\,df_t\,dg_t \mid \mathcal{F}_t] \;=\; \sigma_{f,t}\sigma_{g,t} f_t g_t\,dt\;.$$

This makes the correlation coefficient between df_t and dg_t equal to one:

$$\mathrm{corr}(df_t, dg_t \mid \mathcal{F}_t) \;=\; \frac{\mathrm{cov}(df_t, dg_t \mid \mathcal{F}_t)}{\sqrt{\mathrm{var}(df_t \mid \mathcal{F}_t)\,\mathrm{var}(dg_t \mid \mathcal{F}_t)}} \;=\; 1 \; .$$

As an aside, note that you can state the hedging argument above as saying that if $\mathrm{var}(d\Pi_t \mid \mathcal{F}_t) = 0$, then $d\Pi_t = r_t \Pi_t dt$. To summarize, reweighting using the function m_t in the Girsanov formula changes the expected return of every traded asset as (11) and (12). The market price is also changed

$$\lambda_t \to \lambda_t \,+\, m_t \; , \tag{13}$$

as you can see by using (11) in (8). A reweighting can change the market price of risk, but it cannot change the fact that it is the same for every traded asset.

The formulas for bond prices are derived by making quotients of traded assets into martingales through rewieghting. Recall that a diffusion X_t is a martingale if $E[dX_t \mid \mathcal{F}_t] = 0$. To figure out the expected change of quotients, we use the Taylor expansion

$$\frac{1}{1+\epsilon} \;=\; 1 - \epsilon + \epsilon^2 + O\left(\epsilon^3\right) \; .$$

Continuing,

$$\frac{1}{x+\epsilon} \;=\; \frac{1}{x\left(1+\frac{\epsilon}{x}\right)} \;=\; \frac{1}{x}\,\frac{1}{1+\frac{\epsilon}{x}} \;=\; \frac{1}{x}\left(1 - \frac{\epsilon}{x} + \frac{\epsilon^2}{x^2} + O\left(\epsilon^3\right)\right) \;=\; \frac{1}{x} - \frac{\epsilon}{x^2} + \frac{\epsilon^2}{x^3} + O\left(\epsilon^3\right) \; .$$

Applying this to the quotient gives (dropping terms like $(dg_t)^3$ as usual)

$$\begin{aligned} d\left(\frac{f_t}{g_t}\right) &= \frac{f_t + df_t}{g_t + dg_t} - \frac{f_t}{g_t} \\ &= (f_t + df_t)\left(\frac{1}{g_t} - \frac{dg_t}{g_t^2} + \frac{(dg_t)^2}{g_t^3}\right) - \frac{f_t}{g_t} \\ &= \left(\frac{df_t}{g_t} - \frac{f_t\, dg_t}{g_t^2}\right) + \left(\frac{-df_t\, dg_t}{g_t^2} + \frac{f_t\,(dg_t)^2}{g_t^3}\right) \; . \end{aligned}$$

The formulas (6) and (7) imply that $df_t\, dg_t \;=\; f_t g_t \sigma_{f,t}\sigma_{g,t}\, dt$, and the above simplifies to

$$d\left(\frac{f_t}{g_t}\right) \;=\; \frac{f_t}{g_t}\left(\mu_{f,t} \,-\, \mu_{g,t} \,-\, \sigma_{f,t}\sigma_{g,t} \,+\, \sigma_{g,t}^2\right) \;+\; \frac{f_t}{g_t}\left(\sigma_{f,t} \,-\, \sigma_{g,t}\right) dW_t \; .$$

Therefore, $\frac{f_t}{g_t}$ is a martingale in a given weighing (13) if, in that weighting,

$$\mu_{f,t} \,-\, \mu_{g,t} \,-\, \sigma_{f,t}\sigma_{g,t} \,+\, \sigma_{g,t}^2 \;=\; 0 \; . \tag{14}$$

We start by substituting (8) and (13), which puts this condition in the form

$$r + \lambda_t \sigma_{f,t} + \sigma_{f,t} m_t - r - \lambda_t \sigma_{g,t} - \sigma_{g,t} m_t \,-\, \sigma_{f,t}\sigma_{g,t} \,+\, \sigma_{g,t}^2 \;=\; 0 \; ,$$

which simplifies to

$$\lambda_t + m_t - \sigma_{g,t} = 0 .$$

This gives the following conclusion: the reweighting that makes f_t/g_t into a martingale is the reweighting that makes $\lambda_t = \sigma_{g,t}$ – the market price of risk is equal to the volatility of g. Notice that this does not depend on which f you use. If f_t/g_t is a martingale, and if h_t is another traded asset in this one factor world, then h_t/g_t also is a martingale.

The denominator, g_t, is called the *numeraire*. The ratio f_t/g_t is the size of f_t measured in units of g_t. The numeraire is the standard unit used to measure the sizes of other things. The reweighted measure in which f_t/g_t is a martingale is the *martingale measure* for numeraire g_t. Any traded asset can be used as the numeraire. Of course, there is a reweighting in which $\lambda_t = 0$. This is the *risk neutral* measure, as before.

3 Money market as numeraire

The money market fund is the totally liquid fund that is always invested at the risk free rate. You can think of the money market fund as being invested in the overnight rate every day. This will earn the short rate in each time interval dt. That means that

$$dM_t = r_t M_t \, dt . \tag{15}$$

We may as well assume $M_0 = 1$. We can integrate the ordinary differential equation (15), and use the initial condition, to get

$$M_t = e^{\int_0^t r_s \, ds} . \tag{16}$$

If you take the Ito differential of (16), you get (15) with no Ito term because $(dM_t)^2 = r_t^2 M_t^2 (dt)^2 = 0$. Of course, M_t is a tradable asset, as one can invest as much money as one likes in the short rate and add or remove funds at will. This is the definition of tradable asset. Because $\sigma_{t,M} = 0$, the martingale measure for M_t, the reweighting that makes f_t/M_t a martingale, is the risk neutral measure: $\lambda = \sigma_{M,t}$.

Consider a traded zero coupon bond that matures at time T. Before it matures, its price is $B(t,T)$. In the money market, risk neutral, martingale measure,

$$\frac{B(t,T)}{M_t}$$

is a martingale, when considered as a function of t with T being a fixed parameter. Therefore (and this is the main point)

$$\frac{B(0,T)}{M_0} = E_{RN}\left[\frac{B(T,T)}{M_T}\right] .$$

With the information available at time $t = 0$, the left side is known, which explains why there is no expectation on the left. Also, $M_0 = 1$ is the initial value of the money market fund. Finally $B(T,T) = 1$. Altogether, we get a risk neutral formula for the bond price today, using (16):

$$B(0,T) \;=\; E_{RN}\left[e^{-\int_0^T r_t dt} \right] \; . \tag{17}$$

4 Models of the short rate

Using the pricing formula (17), any model of the short rate gives a prediction of the shape of the yield curve. These models of r_r are different from the stock price (equity) models we used before. There we derived the risk neutral model from the real world model using the arbitrage argument. Here we directly model the risk neutral measure.

Generally speaking there are two kinds of one factor short rate models, *equilibrium* models and *no arbitrage* models. The term *equilibrium model* comes from the fact that these models predict a long term equilibrium model for the probability distribution of r_t. This is possible because, unlike stocks, there is no reason the short rate cannot be *mean reverting*, tending to a natural long term value. These models have a small number of parameters that can be calibrated by using a known yield curve. They have the drawback that they cannot fit any given yield curve. The more complex equilibrium models have more fitting parameters and can fit a wider range of yield curve shapes.

No arbitrage models contain whole functions of time as parameters. For this reason, they can fit a more or less arbitrary yield curve shape. This is called "no arbitrage" because if you offered a bond for sale at a price different from the yield curve, that would create an arbitrage opportunity for your customers.

The simplest equilibrium model that makes any sense is the *Vasicek* model

$$dr_t \;=\; a(\overline{r} - r_t)dt \,+\, \sigma dW \; . \tag{18}$$

The parameters are a, the *mean reversion rate*, $\overline{r}$, the *equilibrium interest rate*, and σ, the volatility of the short rate. Mathematicians call the diffusion process that satisfies

$$dX_t \;=\; -aX_t dt \,+\, \sigma dW_t$$

the *Ornstein Uhlenbeck* process. The substitution $X_t = r_t - \overline{r}$ turns the Vasicek model into the Ornstein Uhlenbeck process. The solution is given by

$$X_t \;=\; e^{-at}X_0 \,+\, \sigma \int_0^t e^{-a(t-s)}\, dW_s \; .$$

In terms of the short rate, this is

$$r_t \;=\; \overline{r} \,+\, e^{-at}\left(r_0 - \overline{r}\right) \,+\, \sigma \int_0^t e^{-a(t-s)}\, dW_s \; . \tag{19}$$

This means that r_t is a gaussian random variable with mean

$$E\left[r_t\right] \;=\; \overline{r} \,+\, e^{-at}\left(r_0 - \overline{r}\right) \; ,$$

and variance

$$\mathrm{var}\left[r_t\right] \;=\; \sigma^2 \int_0^t e^{-2a(t-s)}\, ds \;=\; \frac{\sigma^2}{2a}\left(1 - e^{-2at}\right)\; .$$

Clearly, mean and variance of r_r converge to finite values $\overline{r}$ and $\sigma^2/2a$ as $t \to \infty$. This is the limiting equilibrium distribution of r_t, which confirms the statement that it is an equilibrium model.

Moreover, it is possible to evaluate the expectation (17) explicitly. One way to see this is that the exponent $Z = -\int_0^T r_t dt$ is a Gaussian random variable as well. We know that $E\left[e^Z\right] = e^{\mu_Z + \sigma_Z^2/2}$. Integrating (19) allows us to evaluate $\mu_Z = -E\left[\int_0^T r_t dt\right]$ and $\sigma_Z^2 = \mathrm{var}\left(\int_0^T r_t dt\right)$ explicitly. The result is that when T is large,

$$B(0,T) \;\approx\; e^{-\left(\overline{r} - \frac{\sigma^2}{2a^2}\right)T}\; , \tag{20}$$

You get the effective yield, $\overline{Y}_T$, (these are the numbers plotted in the *yield curve*) for $B(0,T)$ with

$$e^{-\overline{Y}_T T} \;=\; B(0,T)\; ,$$

This, together with (20) gives the long term yield as

$$\overline{Y}_T \;\approx\; \overline{r} - \frac{\sigma^2}{2a^2}$$

Note that although the short rate mean reverts to $\overline{r}$, the long term yield is less than the average short rate by an amount that depends on the volatility and the rate of mean reversion.

There are several criticisms of the simple Vasicek model. One is that it has a limited range of shapes of the yield curve. Another is that it allows the interest rate sometimes to be negative. The *Cox Ingersoll Ross* model, called *CIR* most of the time, fixes the latter. It is

$$dr_t \;=\; a\left(\overline{r} - r_t\right)\, dt \;+\; \sqrt{r_t}\sigma\, dW_t\; . \tag{21}$$

The factor $\sqrt{r}$ in front of the noise effectively turns off the noise as $r \to 0$, which means that the noise will not push r into negative territory. The yield curve shapes for the CIR model are similar to those of the Vasicek model, but the formulas are more complicated.