Estimating Correlated Diffusions

Robert A. Jones*

November 1999

Abstract

The paper derives and tests maximum likelihood parameter estimators for symmetrically correlated Weiner processes observed at discrete intervals. Such processes arise when pricing and determining Value-at-Risk for portfolio derivatives. Cases of driftless and mean-reverting state variables are considered. The procedure is applied to jointly evolving credit qualities in a portfolio of bonds.

*Simon Fraser University and Wells Fargo Bank. Address: Economics Department, Simon Fraser University, Burnaby, B.C. V5A 1S6. tel: 604-291-3367. email: rjones@sfu.ca

The securitization of and creation of credit derivatives based on large portfolios in the banking industry is expanding at a rapid pace. The theoretical value of such contracts depends critically on the degree of comovement of default risk among borrowers, or other sources of correlation in security returns, within the portfolio. Similarly, intra-portfolio correlations are a primary determinant of Value-at-Risk in large portfolios and thus concern management, rating agencies and regulators. Tractable correlation structures and matching estimation methods are essential for the business of banking. This paper provides maximum likelihood estimators for the parameters of symmetric, correlated Weiner processes observed at discrete intervals. Situations are considered where the underlying processes have either zero drift or are mean-reverting. This is a natural starting point for the problem of portfolio derivatives.

The paper is organized as follows. Section I sets out candidate correlation structures in continuous time. Section II obtains the discrete time processes that would correspond to feasibly observable data. Section III derives maximum likelihood estimators for the case of zero drift processes. Section IV repeats this for mean-reverting processes. Section V describes how asymptotic standard errors can be obtained. Section VI performs simulation testing, both to verify large sample and to explore small sample properties of the estimators. Finally section VII provides an illustrative application to jointly evolving (latent) credit quality of publicly traded firms' debt. An Appendix extends the estimators to situations with incomplete or partial observations.

I. Candidate correlation structures

Suppose we have a number of state variables $x_i(t)$ whose evolution in continuous time can be described by stochastic differential equations

$$dx_i = \alpha_i(x,t)\,dt + \sigma_i(x,t)\,dz_i \qquad i = 1,\ldots,n \tag{1}$$

in which dz_i are increments in standard Weiner processes with correlations $\rho_{ij}(x,t)$ between them. For example, x_i might describe the credit quality of a given borrower or, price of a given security, within a portfolio.

The difficulty with this modestly general specification is that, in situations of interest, n may be large and specific information about individual i's either unavailable or too costly to warrant acquisition. Moreover historical data available is likely about specific firms or individuals that are distinct from the group relevant for an application at hand. Let us thus assume that data has already been grouped so that the drift and volatility functions, α_i and σ_i, are the same for all i both in the historical sample and in some current application. Similarly, we must have simple structures for the correlation coefficients ρ_{ij}, both so that they may be reliably estimated with the limited and incomplete data likely to be available, and so that parameter estimates may be applied to borrowers or securities viewed as being of the same generic type, but for which no history is available.

With these considerations in mind, the particular structure we examine in this paper is the linear, constant volatility, single common factor case:

$$dx_i = \kappa(\mu - x_i)\,dt + \sigma(\rho^{1/2}\,dz_0 + (1-\rho)^{1/2}\,dz_i) \tag{2}$$

in which the $z_i(t)$, $i = 0 \ldots n$ are *independent* standard Weiner processes and the parameters $\kappa, \mu, \sigma, 0 \leq \rho \leq 1$ are constants.[1] This structure includes the cases of zero, constant (as a

[1] The process extends to situations of $-1/(n-1) \leq \rho < 0$ by defining $dz_0 \equiv \sum_1^n dz_i$ and changing

Table 1: Alternative correlation structures

dx_i	Total volatility	Correlation	No. of parameters
$\sigma(\rho^{1/2}\,dz_0 + (1-\rho)^{1/2}\,dz_i)$	σ	ρ	2
$\sigma_i(\rho^{1/2}\,dz_0 + (1-\rho)^{1/2}\,dz_i)$	σ_i	ρ	$n+1$
$\sigma(\rho_i^{1/2}\,dz_0 + (1-\rho_i)^{1/2}\,dz_i)$	σ	$(\rho_i\rho_j)^{1/2}$	$n+1$
$\sigma_i(\rho_i^{1/2}\,dz_0 + (1-\rho_i)^{1/2}\,dz_i)$	σ_i	$(\rho_i\rho_j)^{1/2}$	$2n$

limiting case), and mean-reverting drift. z_0 has the interpretation of a common factor giving rise to correlation between the movements of the various x_i. The resulting correlation matrix has 1's on the diagonal and ρ in all off-diagonal locations. Our problem is to estimate the four fixed parameters from a time series of observations at discrete intervals of the x_i.

For comparison, alternative simple correlation structures, ranked in order of number of parameters to estimate, are suggested in Table I. (drift terms suppressed). The last case is equivalent to standard factor analysis with a single common factor. Note, however, that all but the first case would require specific further information about a borrower/security not in the estimation group to permit application of results to another group. Thus only the first is considered here.

II. Discrete time likelihood function

Let the n-vector $x(t) \equiv (x_i)$ be observed at T equally-spaced intervals of length h. Assume x follows a constant coefficient, linear process in continuous time

$$dx = (Ax + b)\,dt + dz \qquad \text{with} \qquad \mathbf{E}(\,dz\,dz') = \Omega\,dt \tag{3}$$

A and b are respectively a $n \times n$ matrix and a column n-vector of constants. Following Wymer (1972), the exact discrete time process for x is

$$x(t+h) = e^{hA}x(t) + A^{-1}[e^{hA} - I]b + \eta_t \qquad \text{where} \qquad \eta_t \sim \mathrm{N}(0, \int_0^h e^{\tau A}\Omega e^{\tau A'}\,d\tau) \tag{4}$$

I denotes the $n \times n$ identity matrix. I.e., the distribution of $x(t+h)$ conditional on $x(t)$ is joint normal. The expression e^A is defined as Ve^DV^{-1}, where V is a matrix whose columns are the eigenvectors of A, and e^D is a diagonal matrix with elements e^{c_i}, where the c_i are the corresponding eigenvalues of A. Note that the eigenvectors of hA are the same as for A but with corresponding eigenvalues of hc_i.

For the process of equation (2), these components are

$$A = -\kappa I \qquad A^{-1} = -\frac{1}{\kappa}I \qquad b = \kappa\mu\mathrm{e} \qquad \Omega = \sigma^2[(1-\rho)I + \rho\mathrm{e}\mathrm{e}'] \tag{5}$$

in which e denotes a column vector of 1's. Observing that A has n eigenvalues all equal to $-\kappa$ with eigenvectors being the n unit vectors e_i (ith element 1 and the rest 0), the covariance

the random term in (2) to $\sigma((1-\rho)^{1/2}\,dz_i - ((1-\rho)^{1/2} - (1+n\rho-\rho)^{1/2})/n\;dz_0)$. For more negative ρ, the covariance matrix is not positive-definite.

matrix of $x(t+h)$ is obtained:

$$\int_0^h e^{\tau A}\Omega e^{\tau A'}\,d\tau = \int_0^h e^{-s\kappa} I\Omega I' e^{-s\kappa'}\,d\tau = \Omega \int_0^h e^{-2s\kappa}\,d\tau = \frac{1-e^{-2h\kappa}}{2\kappa}\,\Omega \tag{6}$$

Substituting these relations into equation (4) and rearranging,

$$x(t+h) - \underbrace{e^{-h\kappa}}_{a} x(t) - \underbrace{(1-e^{-h\kappa})\mu}_{b}\,\mathrm{e} \;\sim\; \mathrm{N}(\,0\,,\, \underbrace{\frac{(1-e^{-2h\kappa})}{2\kappa}}_{s}\sigma^2[(1-\rho)I + \rho\mathrm{ee}']\,) \tag{7}$$

It is this expression that forms the basis for the likelihood function.

The likelihood function will be expressed, and estimation conducted, in terms of parameters a, b, s as defined in equation (7). The continuous time parameters are then retrieved from the one-to-one relationships

$$\kappa = -\frac{1}{h}\ln a \qquad \mu = \frac{b}{1-a} \qquad \sigma^2 = \frac{2s\ln a}{h(a^2-1)} \qquad \rho = \rho \tag{8}$$

The zero drift case obtains by setting $a = 1$ and $b = 0$. The constant drift case obtains by setting $a = 1$, estimating b, and noting that the continuous time drift rate is simply b/h.

Now suppose that we have observations on the n state variables at equally spaced times t_j, $j = 1 \ldots T+1$. Define the n-vector

$$y_j \equiv x(t_{j+1}) - ax(t_j) - b\mathrm{e} \qquad\qquad j = 1 \ldots T \tag{9}$$

Being joint normally distributed, and independent because the x process is Markov and the time intervals do not overlap, the likelihood function for these observations is

$$L = \prod_{j=1}^{T} \frac{1}{(2\pi)^{n/2}|\tilde{\Omega}|^{1/2}} e^{-y_j'\tilde{\Omega}^{-1}y_j/2} \tag{10}$$

where from equation (7)

$$\tilde{\Omega} \equiv s[(1-\rho)I + \rho\mathrm{ee}'] \tag{11}$$

Maximizing L is equivalent to maximizing Λ, defined as

$$\Lambda \;\equiv\; 2\ln L \;=\; -nT\ln(2\pi) - T\ln|\tilde{\Omega}| - \textstyle\sum_{j=1}^{T} y_j'\tilde{\Omega}^{-1}y_j \tag{12}$$

We are almost there. One may verify that that $\tilde{\Omega}$ satisfies the following:[2]

$$|\tilde{\Omega}| \;=\; s^n(1-\rho)^{n-1}(1+n\rho-\rho) \qquad\qquad \tilde{\Omega}^{-1} \;=\; \frac{1}{(1-\rho)s}[I - \frac{\rho\mathrm{ee}'}{1+n\rho-\rho}] \tag{13}$$

Substituting into Λ gives

$$\Lambda = -nT\ln(2\pi) - nT\ln s - (n-1)T\ln(1-\rho) - T\ln(1+n\rho-\rho) - \frac{\sum_j y_j'y_j}{s(1-\rho)} + \frac{\rho\sum_j y_j'\mathrm{ee}'y_j}{s(1-\rho)(1+n\rho-\rho)} \tag{14}$$

[2]One may also verify that $\tilde{\Omega}$ has unique largest eigenvalue of $(1+n\rho-\rho)s$ with eigenvector e, and $n-1$ eigenvalues of value $(1-\rho)s$ with (non-unique) eigenvectors $-\mathrm{e}_i + (\mathrm{e} + \sqrt{n}\mathrm{e}_1)/(n+\sqrt{n})$, $i = 2 \ldots n$.

The final step is to substitute for y_j from (9). Letting x_j denote $x(t_j)$ and $\tilde{x}_j$ denote $x(t_{j+1})$, yields the (two times) log-likelihood in terms of the model parameters and moments of the data. Summations are understood to be over $j = 1 \ldots T$.

$$\begin{aligned}
\Lambda \;=\; & -nT\ln(2\pi) - nT\ln s - (n-1)T\ln(1-\rho) - T\ln(1+n\rho-\rho) \\
& - \frac{nTb^2 + \Sigma\tilde{x}_j'\tilde{x}_j - 2a\Sigma x_j'\tilde{x}_j + a^2\Sigma x_j'x_j + 2ab\Sigma x_j'\mathrm{e} - 2b\Sigma\tilde{x}_j'\mathrm{e}}{s(1-\rho)} \\
& + \frac{\rho(Tn^2b^2 + \Sigma\tilde{x}_j'\mathrm{ee}'\tilde{x}_j - 2a\Sigma x_j'\mathrm{ee}'\tilde{x}_j + a^2\Sigma x_j\mathrm{ee}'x_j + 2nab\Sigma x_j'\mathrm{e} - 2nb\Sigma\tilde{x}_j'\mathrm{e})}{s(1-\rho)(1+n\rho-\rho)}
\end{aligned} \tag{15}$$

III. Maximum likelihood estimators: zero drift

The section specializes to the case of zero drift by setting $a = 1$ and $b = 0$. The likelihood function reduces to

$$\begin{aligned}
\Lambda \;=\; & -nT\ln(2\pi) - nT\ln s - (n-1)T\ln(1-\rho) - T\ln(1+n\rho-\rho) \\
& - \frac{\sum(\tilde{x}_j - x_j)'(\tilde{x}_j - x_j)}{s(1-\rho)} + \frac{\rho\sum(\tilde{x}_j - x_j)'\mathrm{ee}'(\tilde{x}_j - x_j)}{s(1-\rho)(1+n\rho-\rho)}
\end{aligned} \tag{16}$$

Taking partial derivatives with respect to s and ρ, equating to 0 then solving, gives explicit maximum likelihood estimators:

$$\begin{aligned}
\hat{s} \;&=\; \frac{\sum(\tilde{x}_j - x_j)'(\tilde{x}_j - x_j)}{nT} \\
&\equiv\; \frac{1}{nT}\textstyle\sum_{j=1}^{T}\sum_{i=1}^{n}(\tilde{x}_{ij} - x_{ij})^2
\end{aligned} \tag{17}$$

$$\begin{aligned}
\hat{\rho} \;&=\; \frac{\sum(\tilde{x}_j - x_j)'[\mathrm{ee}' - I](\tilde{x}_j - x_j)}{(n-1)\sum(\tilde{x}_j - x_j)'(\tilde{x}_j - x_j)} \\
&\equiv\; -\frac{1}{n-1} + \frac{1}{(n-1)nT\hat{s}}\textstyle\sum_{j=1}^{T}(\sum_{i=1}^{n}(\tilde{x}_{ij} - x_{ij}))^2
\end{aligned} \tag{18}$$

$$\Lambda^* \;=\; -T\left((\ln 2\pi + \ln\hat{s} + 1)n + (n-1)\ln(1-\hat{\rho}) + \ln(1+n\hat{\rho}-\hat{\rho})\right) \tag{19}$$

The continuous time volatility estimate is related to $\hat{s}$ by

$$\hat{\sigma} \;=\; (\hat{s}/h)^{1/2} \tag{20}$$

An unbiased estimate of the common component of the state change over the j^{th} observation interval, $\sigma\rho^{1/2}(z_0(t_j + h) - z_0(t_j))$, is simply the average x change

$$\hat{\epsilon}_j \;=\; \frac{1}{n}\textstyle\sum_{i=1}^{n}(\tilde{x}_{ij} - x_{ij}) \tag{21}$$

This will be a noisy estimate, with error variance converging to 0 only as the number of diffusions n in the cross-section goes to infinity.

Making use of the true covariance matrix from (7), one can verify that $\mathbf{E}(\hat{s}) = \sigma^2$ and $\mathbf{E}(\hat{\rho}\hat{s}) = \rho\sigma^2$. The maximum likelihood estimates of the instantaneous variance/covariance of the x-process are thus unbiased. However the maximum likelihood estimates of the instantaneous volatility and correlation (σ and ρ), being nonlinear functions of the estimated variance/covariance, will be biased in finite samples.

IV. Maximum likelihood estimators: mean reversion

For the general mean-reverting process case, the likelihood function is as in (15). However the ML estimators are too unwieldy to present in their entirety. Indeed, we resort partially to numerical optimization of Λ as described below.

We proceed as follows. First, maximize Λ with respect to a, b, s for given ρ by setting the partial derivatives with respect to those variables equal to 0 and solving for their values. This gives

$$a = \frac{nT(1+n\rho-\rho)\Sigma x_j'\tilde{x}_j - \rho nT\Sigma x_j'\mathrm{ee}'\tilde{x}_j - (1-\rho)\Sigma x_j'\mathrm{e}\Sigma\tilde{x}_j'\mathrm{e}}{nT(1+n\rho-\rho)\Sigma x_j'x_j - \rho nT\Sigma x_j'\mathrm{ee}'x_j - (1-\rho)\Sigma x_j'\mathrm{e}\Sigma x_j'\mathrm{e}} \tag{22}$$

$$b = \frac{\Sigma\tilde{x}_j'\mathrm{e} - a\Sigma x_j'\mathrm{e}}{nT} \tag{23}$$

$$s = \frac{1}{nT}\Sigma(\tilde{x}_j - ax_j - b\mathrm{e})'(\tilde{x}_j - ax_j - b\mathrm{e}) \tag{24}$$

Note that s is expressed above in terms of the expressions for a, b. Substitution of these into Λ gives a concentrated likelihood function $\Lambda^*(\rho)$. We maximize this numerically with respect to ρ to obtain $\hat{\rho}$, substituting the outcome into (22) to (24) to get $\hat{a}, \hat{b}, \hat{s}$.[3]

An estimate of the state change attributable to movement in the common factor over the j^{th} observation interval, $(\rho s/h)^{1/2}(z_0(t_j+h) - z_0(t_j))$, is the average x change with estimated drift removed[4]

$$\hat{\epsilon}_j = \frac{1}{n}\sum_{i=1}^{n}(\tilde{x}_{ij} - \hat{a}x_{ij} - \hat{b}) \tag{25}$$

An estimate of the change in z_0—for comparison, say, with changes in other factors—is $\hat{\epsilon}_j/(\hat{\rho}\hat{s}/h)^{1/2}$. Note that this will be biased in small samples because only estimated ρ, s values are available. It will also be noisy for arbitrarily large T (but small n) because random comovement of the independent z_i in the same direction will be erroneously attributed to movement in z_0.[5]

V. Distribution of estimated parameters

Under certain regularity conditions, the maximum likelihood estimates of a parameter vector γ are asymptotically distributed around the true γ as follows:[6]

$$T^{1/2}(\hat{\gamma} - \gamma) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \lim(I/T)^{-1}) \tag{26}$$

[3]The first order condition $\partial\Lambda^*/\partial\rho = 0$ is revealed by Maple to be a third order polynomial in ρ, so could in principle be solved analytically. However the polynomial coefficients are very lengthy expressions involving sixth moments of the data. Numerical maximization using the BRENT subroutine from *Numerical Recipes* seemed the more expedient route.

[4]Compared to no mean reversion, a given movement in z_0 has less impact on the states because its effect is diminished by mean reversion between observation dates. I.e., $\rho s/h < \rho\sigma$ for $\kappa > 0$.

[5]The asymptotic ($T \to \infty$) variance of the error in estimating Δz_0 is $(1-\rho)h/\rho n$.

[6]See Dhrymes (1974, p.122) or Judge *et al* (1985, p.178)

in which I denotes the information matrix

$$I = -\mathbf{E}(\frac{\partial^2 \ln L}{\partial\gamma\partial\gamma'}) \tag{27}$$

L is the sample size T likelihood function evaluated at the *true* γ, and the limit in (26) is as T goes to infinity. There are a variety of methods for obtaining an estimate of this matrix. We adopt a method of Berndt, Hall and Hausmann as given in Judge (1985, p.180, eq.5.6.8). Their estimator for $\lim(I(\gamma)/T$ is

$$\frac{1}{T}\left[\sum_{t=1}^{T}\left(\frac{\partial \ln L_t}{\partial\gamma}\right)\left(\frac{\partial \ln L_t}{\partial\gamma}\right)'\right]_{\gamma=\hat{\gamma}} \tag{28}$$

where L_t denotes the probability density of the one-period observation y_t. Thus, for each observation date separately, we determine numerically the partial derivatives of the log-likelihood with respect to the four parameters $\gamma \equiv (\sigma, \rho, \kappa, \mu)'$ evaluated at the maximum likelihood *estimate* $\hat{\gamma}$, and accumulate the outer product of that vector with itself.[7] Standard errors for the parameters are square roots of the corresponding diagonal elements of the inverse of this matrix.

VI. Simulation testing

To test the estimation method and determine the small sample characteristics of the parameter estimates, hypothetical data sets were created by Monte Carlo simulation and `MAXLIKE` applied to them. From a reverse perspective, this also tests the procedure used to simulate multiple correlated diffusions. Such procedures are required to value contingent claims whose payouts depend on the joint evolution of correlated diffusions within a portfolio (e.g., credit derivatives).

Benchmark parameter values used for the mean-reverting case were $\kappa = 1$, $\mu = 5$, $\sigma = 1$, $\rho = .25$ (assume one year time unit). Observation intervals were set at .25 years. For each variation below, 500 simulations/estimations were performed.[8]

To verify large sample performance, simulations of 100 joint diffusions over 100 observation intervals are reported in Table 2. As can be seen, the average of each parameter estimate agrees closely with the true value used to create the data. Furthermore, the Berndt-Hall-Hausmann estimate of the standard errors agrees quite well with the simulation standard deviations, with no obvious bias in either direction.

Table 3 reports on simulation of 10 joint diffusions for progressively shorter numbers of observation intervals from 50 down to 5. Aside from the understandably larger standard errors as the number of observation intervals shrinks, the most notable feature is the progressively larger bias toward 0 in the estimated ρ, mild downward bias in σ, and upward bias in the mean-reversion coefficient κ. This is consistent with the idea that the randomly occuring 'trend'

[7] An alternative estimator tried based on the numerically evaluated matrix of second partials of the log-likelihood (given in Judge as eq. 5.6.7) , behaved in a less satisfactory manner on some data sets.

[8] Uniform random variables generated by `RAN1` of *Numerical Recipes* were converted to normal deviates using their routine `GASDEV`. These were used in turn to generate x changes following equation (7). The initial distribution of the state variable was set at uniform on [0,10]. Alternative initial distributions (all at 5; half each at 0 and 10) resulted in little difference except in very small samples.

Table 2: Large sample simulation

	κ	μ	σ	ρ	Λ
input value	1.0000	5.0000	1.0000	0.2500	
avg. estimate	1.0030	5.0028	0.9985	0.2457	-2.6452e+03
minimum	0.9216	4.7122	0.9427	0.1539	-2.8386e+03
maximum	1.1221	5.2943	1.0552	0.3273	-2.4628e+03
sample st. dev.	0.0267	0.0966	0.0193	0.0291	6.5957e+01
BHH st. dev.	0.0277	0.1039	0.0203	0.0288	

100 observations of 100 diffusions. 500 Monte Carlo trials. Uniform starting distribution on [0,10].

that will be present in any short series of the common factor can be equally (statistically) construed as stronger reversion toward a mean particular to that sample. A second notable feature is the increasing overstatement by the BHH standard error of the true estimation error—by a factor of 50 for $T = 5$. For $T \geq 20$ the overstatement appears modest enough to be ignored.

Table 4 reports on simulation for 50 observation intervals of progressively fewer joint diffusions from 50 down to 2. Here no consistently developing bias appears to show up in the value of any of the parameters. The BHH standard errors slightly overstate the true standard deviations, but the proportional overstatement is slight (sometimes even slight understatement) and appears unconnected with sample size. Note that 2 is the minimum number of diffusions for which the notion of correlation could have meaning.

Table 3: Varying time series length for $N = 10$

	κ	μ	σ	ρ	Λ
input value	1.0000	5.0000	1.0000	0.2500	
average: $T = 50$	1.0186	5.0250	1.0133	0.2392	-2.0264E+02
standard dev.	0.0932	0.1660	0.0403	0.0502	3.3358E+01
BHH st. dev.	0.1127	0.1703	0.0450	0.0562	
average: $T = 20$	1.0233	5.0085	0.9964	0.2212	-7.5998e+01
standard dev.	0.1052	0.2611	0.0572	0.0784	1.6694e+01
BHH st. dev.	0.1544	0.2816	0.0782	0.0985	
average: $T = 10$	1.0286	5.0793	0.9721	0.2327	-3.0972e+01
standard dev.	0.1137	0.3470	0.0794	0.1053	1.2544e+01
BHH st. dev.	0.2220	0.5270	0.1514	0.2013	
average: $T = 8$	1.0199	5.0678	0.9660	0.2088	-2.4899e+01
standard dev.	0.1154	0.3832	0.0918	0.1232	1.1745e+01
BHH st. dev.	0.2667	0.6535	0.2000	0.2753	
average: $T = 6$	1.0435	5.1334	0.9652	0.1892	-1.8685e+01
standard dev.	0.1207	0.4361	0.1106	0.1321	1.0929e+01
BHH st. dev.	0.4861	2.3931	0.4010	0.6070	
average: $T = 5$	1.0489	5.1747	0.9547	0.1365	-1.5684e+01
standard dev.	0.1309	0.5038	0.1217	0.1392	1.1105e+01
BHH st. dev.	5.7953	14.6454	8.6073	6.7456	

Observations of 10 diffusions. 500 Monte Carlo trials. Uniform starting distribution. T equals number of observation intervals.

Table 4: Varying number of diffusions for $T = 50$

	κ	μ	σ	ρ	Λ
input value	1.0000	5.0000	1.0000	0.2500	
average: $N = 50$	1.0025	5.0101	1.0016	0.2433	-7.4240E+02
standard dev.	0.0468	0.1406	0.0296	0.0398	7.5991E+01
BHH st. dev.	0.0546	0.1533	0.0321	0.0440	
average: $N = 20$	1.0065	5.0097	1.0086	0.2510	-3.3769E+02
standard dev.	0.0648	0.1702	0.0331	0.0423	4.7403E+01
BHH st. dev.	0.0873	0.1632	0.0391	0.0508	
average: $N = 10$	1.0186	5.0250	1.0133	0.2392	-2.0264E+02
standard dev.	0.0932	0.1660	0.0403	0.0502	3.3358E+01
BHH st. dev.	0.1127	0.1703	0.0450	0.0562	
average: $N = 5$	1.0323	5.0346	1.0130	0.2333	-1.1241E+02
standard dev.	0.1217	0.1743	0.0479	0.0660	2.1210E+01
BHH st. dev.	0.1532	0.1860	0.0589	0.0729	
average: $N = 3$	1.0693	5.0241	0.9988	0.2222	-6.6727E+01
standard dev.	0.1564	0.1951	0.0563	0.0943	1.5212E+01
BHH st. dev.	0.1869	0.1991	0.0692	0.1002	
average: $N = 2$	1.0464	5.0673	0.9831	0.2515	-4.2334E+01
standard dev.	0.1633	0.2360	0.0790	0.1386	1.5160E+01
BHH st. dev.	0.2059	0.2304	0.0847	0.1474	

$T = 50$ observation intervals. 500 Monte Carlo trials. Uniform starting distribution. N equals number of diffusions observed.

Table 5: Credit series properties

Average x	Minimum x	Maximum x	Drift	Volatility	Autocorrel
7.45	3.13	8.87	.08	.552	-.042

VII. Application to credit state histories

This final section applies the procedure to a time series of credit states for 104 U.S. firms with publicly traded debt. The latent credit states x_{it} are continuous variables with theoretical range $[0, \infty)$ that are assumed to follow mixed jump-diffusion processes over time. Default is associated with diffusing or jumping to a credit state less than or equal to 0. The credit state of a given firm on a given date is not observed directly, but is inferred from the process parameters and the current market prices of the firm's traded bonds in the context of an arbitrage-free valuation model.[9]

Month-end observations of the firms' outstanding publicly traded debt was obtained for the period May, 1993, to December, 1997. The process of inferring credit states is not described here; rather the output of that process is taken as the data for the current exercise.[10] We thus start with a time series of 54 monthly observations of inferred credit states for each of the 104 firms. Our maintained hypothesis is that the credit states of all firms follow mean-reverting correlated diffusions with identical parameters, as assumed by the estimation procedure. Our problem is to estimate the common values of $\kappa, \mu, \sigma, \rho$.

Summary descriptive statistics for the series are given in Table 5. The drift and volatility are expressed as annual rates. The range of credit qualities corresponds roughly to AAA to B rated bonds. None of the issuers in the sample went bankrupt during the period. As can be seen, credit quality displayed a slight upward drift over the sample period, annual standard deviation of change of approximately half a unit, and slight negative month-to-month serial correlation of changes for any given firm.

Table 6 presents maximum likelihood estimates of the process parameters under assumptions of zero drift and of mean-reversion respectively. Included are the estimated standard errors and correlation matrix of the parameter estimates using the BHH method described earlier. First order autocorrelation coefficients of the estimated common factor series (supposed to be 0 under the assumed specification) were -.106 and -.121 respectively. The very high correlation between the estimated ρ and σ is likely a consequence of these two parameters predominantly appearing as a package $\sigma(1 - \rho)$ in the likelihood function, though it may indicate a weakness of the BHH estimate of the information matrix on which this correlation is based.

The parameter estimates indicate noticable mean-reversion in individual firms' credit quality over the sample and correlation in credit quality changes across firms. The difference of

[9] Appropriate account must be taken of the term structure of default-free bonds at the same time.

[10] The credit evolution model and its estimation are set out in Chau and Jones (1999 in progress). Loosely speaking, credit state here corresponds to the observed market yield spread relative to U.S. Treasuries for a hypothetical standard maturity bond.

Table 6: Estimation results for credit quality series

	κ	μ	σ	ρ	Λ^*
estimate			.5604	0.2567	9704
BHH st. dev.			.0109	0.0292	
correlation matrix			1.000		
			.995	1.000	
estimate	.2906	7.736	.5609	0.2544	9809
BHH st. dev.	.0116	0.545	.0120	0.0325	
correlation matrix	1.000				
	.028	1.000			
	-.080	.490	1.000		
	-.116	.495	.996	1.000	

149 between Λ^* under the two drift assumptions, which should be asymptotically distributed $\chi^2(2)$ if there is no drift, strongly rejects a hypothesis of zero drift against the mean-reversion alternative. However its quantitative importance would depend on the time horizon of the application: With credit quality 2 units away from the target mean of 7.7, for example, the contribution of expected annual drift of $(.29)(2) = .58$ is of the same order of magnitude as the annual volatility of .56 . For horizons well under a year, the volatility will tend to dominate the drift, and the latter might plausibly be ignored; however for multiyear horizons, it will clearly be important. The ρ estimate of .254 means that 25.4% of the overall variance of the monthly changes can be attributed to movements of a common factor that equally affects all firms.

To get a sense of whether imposing the assumption of symmetric correlation is appropriate, conventional factor analysis was applied to the monthly credit state changes. Expected drift, based on the estimates of κ and μ above, was first removed. The remaining random components of changes had a neglibible annualized drift of -.002, volatility of .546, and slight negative autocorrelation of -.041. Table 7 summarizes results for the first three factors in order of importance. Since factor analysis permits different weights (loadings) on the common factors for each variable, the estimated loadings vary by firm in the sample. Reported for each factor are the minimum and maximum loadings in addition to the average and standard deviation across the 104 firms.

The results of Table 7 reveal that a single common factor, optimally weighted for each

Table 7: Factor analysis of credit state changes

	Var explained	Min load	Max load	Avg load	Stdev load
1st factor	.289	-.040	.169	.093	.032
2nd factor	.087	-.121	.774	-.001	.098
3rd factor	.074	-.409	.603	-.006	.098

firm, could explain 28.9% of the overall fluctuation in credit quality—a modest increase over the 25.4% when equal loading was imposed as above. Though there is some dispersion, the weights were mostly of the same sign and moderately clustered around their average value of .093. In contrast, the next two most significant factors had incremental explanatory value of just 8.7% and 7.4% of overall variance respectively, with much greater dispersion across firms in their weights. Inspection of the individual loadings reveals both of these factors to be heavily weighted by just one or two firms, with negligible weight on them by most others. I.e., they appear to embody more firm-specific fluctuation. We conclude from this that imposing symmetric correlation is a useful and not-unreasonable approximation to describing credit state evolution across firms in the sample, but that assuming common idiosyncratic volatility is somewhat at odds with the data.

VIII. Conclusion

This paper has suggested a specification for correlated diffusions more basic than factor analysis with widespread potential application in the financial industry. It recognizes the reality that inference and valuation must often be based on limited observation of generic portfolios with changing and anonymous constituents. It also recognizes that the dominant valuation frameworks require knowledge of underlying continuous time processes, but the statistician must work with discrete time observation. Within these constraints, we have provided, tested and hopefully displayed the applicability of maximum likelihood procedures.

References

Berndt, E. K., B. H. Hall and J. A. Hausman, "Estimation and Inference in Non-linear Structural Models," *Annals of Economic and Social Measurement*, 3 (1974), 653-65.

Dhrymes, Phoebus J., *Econometrics: Statistical Foundations and Applications*, New York: Springer-Verlag, 1974.

Judge, George. G., W. E. Griffiths, R. C. Hill, H. Lutkepohl and T. C. Lee, *The Theory and Practice of Econometrics*, New York: John Wiley & Sons, 1985.

Lo, Andrew W., "Maximum Likelihood Estimation of Generalized Ito Processes with Discretely Sampled Data," *Journal of Econometric Theory*, 4 (1988), 231-247.

Wymer, C.R., "Econometric Estimation of Stochastic Differential Equation Systems," *Econometrica*, Vol. 40, No. 3 (May 1972), 565-578.

Appendix

A. Partial data series

The section indicates how the estimators extend to situations with partial observation series for some or all of the state variables. For example, in the context of borrower credit states, some borrowers/firms may have no credit history or publicly traded debt at the start of observation, or they may drop out part way through by maturity of their debt or by default. We assume that each state variable, for the period that it is visible, is observed at the same interval h and is synchronized with the other state variables then being observed.

Let n_j denote the number of state variables observed at both time t_j and time t_{j+1}, with $y_j = x_{j+1} - ax_j - b\mathrm{e}$ being the n_j-vector of random parts of the changes as in equation (9). The overall likelihood function is as in equation (10) with n becoming n_j and recognizing that the covariance matrix $\tilde{\Omega}$ is of size n_j. Making this adjustment results in the revised two-times log-likelihood function corresponding to equation (14):

$$\begin{aligned}\Lambda &= -\Sigma_j n_j \ln 2\pi - \Sigma_j (n_j - 1)\ln(1-\rho) - \Sigma_j n_j \ln s - \Sigma_j \ln(1 + n_j\rho - \rho) \\ &\quad - \frac{1}{(1-\rho)s}\Sigma_j y_j' y_j + \frac{\rho}{(1-\rho)s}\Sigma_j \frac{y_j'\mathrm{ee}'y_j}{1+n_j\rho-\rho} \qquad \text{(A.1)}\end{aligned}$$

Let $N \equiv \Sigma_j n_j$ and $g_j \equiv 1/(1+n_j\rho-\rho)$ to simplify notation. Substituting for y_j and collecting terms allows Λ to be written in terms of the parameters and data moments as

$$\begin{aligned}\Lambda &= -N\ln 2\pi - (N-T)\ln(1-\rho) - N\ln s + \Sigma \ln g_j \\ &\quad - \frac{1}{(1-\rho)s}\left[\Sigma \tilde{x}_j'\tilde{x}_j - 2a\,\Sigma x_j'\tilde{x}_j + a^2\,\Sigma x_j' x_j\right] - \frac{1}{s}\left[b^2\,\Sigma g_j - 2b\,\Sigma g_j \tilde{x}_j'\mathrm{e} + 2ab\,\Sigma g_j x_j'\mathrm{e}\right] \\ &\quad + \frac{\rho}{1-\rho}\left[\Sigma g_j \tilde{x}_j'\mathrm{ee}'\tilde{x}_j - 2a\,\Sigma g_j x_j'\mathrm{ee}'\tilde{x}_j + a^2\,\Sigma g_j x_j'\mathrm{ee}'x_j\right] \qquad \text{(A.2)}\end{aligned}$$

Summations are understood to be over $j = 1, T$ and the e vectors of 1's to be of appropriate length n_j.

For given ρ, this may be maximized with respect to a, b, s by setting first partial derivatives equal to 0. Letting $M \equiv \Sigma n_j g_j$, this yields

$$a = \frac{M\,\Sigma x_j'\tilde{x}_j - \rho M\,\Sigma g_j x_j'\mathrm{ee}'\tilde{x}_j - (1-\rho)\,\Sigma g_j x_j'\mathrm{e}\,\Sigma g_j \tilde{x}_j'\mathrm{e}}{M\,\Sigma x_j' x_j - \rho M\,\Sigma g_j x_j'\mathrm{ee}'x_j - (1-\rho)\,\Sigma g_j x_j'\mathrm{e}\,\Sigma g_j \tilde{x}_j'\mathrm{e}} \qquad \text{(A.3)}$$

$$b = \frac{\Sigma g_j \tilde{x}_j'\mathrm{e} - a\,\Sigma g_j x_j'\mathrm{e}}{M} \qquad \text{(A.4)}$$

$$s = \frac{\Sigma \tilde{x}_j'\tilde{x}_j - 2a\,\Sigma x_j'\tilde{x}_j + a^2\,\Sigma x_j' x_j + b^2 N - 2b\,\Sigma \tilde{x}_j'\mathrm{e} + 2ab\,\Sigma x_j'\mathrm{e}}{N} \qquad \text{(A.5)}$$

The concentrated likelihood function $\Lambda^*(\rho)$ is obtained by substituting these values into (A.2). This is then maximized numerically with respect to ρ to get estimates of the four parameters. For the case of no mean reversion, a and b are fixed at 1 and 0 respectively when maximizing with respect to ρ.

N	integer	vector length IT giving diffusions $n_j$each date
IT	integer	no. of observation dates T
H	double	time between observations h
X	double	vector length sum N(J) holding stacked x_{ij}
XH	double	vector length sum N(J) holding stacked $\tilde{x}_{ij}$
IFLAG	integer	flag value MRevert + 2*GenCov + 4*GenZ + 8*Nsame
		MRevert = 0 (0 drift) or 1 (mean-reverting)
		GenCov = 0 (skip COV) or 1 (generate COV)
		GenZ = 0 (skip Z) or 1 (generate Z)
		Nsame = 0 (N differ) or 1 (N(1) applies to all)
K	double	estimated κ
U	double	estimated μ
SIG	double	estimated σ
RHO	double	estimated ρ
LMAX	double	maximized 2×log-likelihood Λ^*
Z	double	vector length IT-1 of estimated common factor changes $z_0(t+h) - z_0(t)$
COV	double	vector length 3 or 10 of lower triangle of parameter covariance matrix (ordered $\sigma, \rho, \kappa, \mu$)

B. Subroutine MAXLIKE

This section briefly describes the Fortran subroutine MAXLIKE that implements the maximum likelihood estimators of the previous sections. The routine is invoked within a Fortran main program by

```
call MAXLIKE( N, IT, H, X, XH, IFLAG, K, U, SIG, RHO, LMAX, Z, COV )
```

The first six arguments are inputs, and not modified by the subroutine; the next five arguments are outputs returned. The final two are optionally returned. Argument descriptions follow.

The subroutine is self-contained (i.e., has no other arguments passed by common from calling routine). Care must be taken that the vectors X, XH are full and store data in the manner interpreted by MAXLIKE. The integer vector N of length IT gives the number of state variables observed on each observation date. The vector X is assumed to be the start-of-period observations stacked on top of each other. I.e., the N(1) x values on date 1, followed without gap by the N(2) values on date 2, etc. Argument XH is the corresponding vector of values of the same state variables time H later in the same order. This routine has one internal size parameter MAXT, which is the maximum number of observation *dates* that can be accomodated, and currently set to 500. Aside from this constraint, arbitarily large data sets are accepted.

In the special case where the set of state variables observed each date is the same, and because of the way arrays are stored in Fortran, the calling program may equivalently dimension `X` as a two-dimensional array of size `(N,IT)`. `X(i,j)` would be the j^{th} observation in time of the i^{th} state variable or borrower. `XH(i,j)` will typically be simply `X(i,j+1)` if movements over the shortest observed time horizon are the basis for estimation. This can be accomplished in Fortran by simply passing `X(1,2)` as argument `XH`. However it may be set so that the observation interval is larger, say `X(i,j+Δj)`, if one wishes to lessen the impact of observation noise or error not specified in the model. Setting `IFLAG` appropriately lets the first element of `N` apply to all dates.

The parameter covariance matrix `COV` and estimated vector of common factor changes `Z` are optionally generated if `IFLAG` are set approriately. Note however that it assumes independent increments in the time series observations and is thus only meaningful if the observation intervals are non-overlapping (i.e., time horizon for estimation is one period). Only the 3 (no mean-reversion) or 10 (with mean-reversion) element lower-triangular portion of this symmetric matrix is returned: i.e., first element is the first diagonal, next two elements are the second row, next three the third, etc.