10ISSN 1183-1057

SIMON FRASER UNIVERSITY

Department of Economics

Working Papers

11-03

"Estimation of Equicorrelated Diffusions from Incomplete Data"

Robert A. Jones
and
Mohammad Zanganeh

October, 2011

SFU Economics

Estimation of Equicorrelated Diffusions from Incomplete Data

Robert A. Jones*

Mohammad Zanganeh†

Simon Fraser University Working Paper

October 2011

Abstract

The paper derives 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 applicable to samples with missing data of any pattern and to high dimensional systems. The estimation procedure is illustrated using a sample of stock prices.

JEL classification: C51, C58, G11, G21

keywords: Maximum likelihood; Equicorrelation; Correlated diffusions; Wiener process; Missing data

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

†Simon Fraser University. email: mzangane@sfu.ca

The securitization of and creation of credit derivatives based on large portfolios in the banking industry has expanded 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, equicorrelated Weiner processes observed at discrete intervals. This is a natural starting point for the problem of portfolio derivatives. Situations are considered where the underlying processes have either zero or constant drift or are mean-reverting.

The estimation procedure developed in this paper addresses a number of practical issues in financial economics. First, it is applicable to the case of partially observed state variables which is quite common in financial applications. 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 may drop out part way through maturity of their debt or default. Estimation methods requiring a complete observation matrix, with no missing data, could fail as the number of missing observations increases. The maximum likelihood estimators presented in this paper have the advantage of being applicable to the samples with missing data of any pattern. Second, since the estimation procedure does not require computation of covariance matrix inverse and determinant, it is readily applicable to high-dimensional systems. Numerical maximization of likelihood functions becomes cumbersome and can breakdown as the number of state variables increases. For example, as reported by Engle and Kelly (2009), the Dynamic Conditional Correlation (DCC) model developed by Engle (2002) has been successfully applied up to just 100 assets. Portfolio size is not a restriction for our estimation procedure. Third, the model provides maximum likelihood estimates of equicorrelation that is widely applied by both practitioners and academics to problems ranging from portfolio selection to risk management and derivatives pricing. Major applications of the equicorrelation structure are briefly reviewed below. Finally, the capability of the estimation procedure to handle incomplete time series provides a natural way of dealing with outliers.

The paper is organized as follows. Section I briefly reviews the major application areas of equicorrelation modeling. Section II sets out candidate correlation structures in continuous time. Section III obtains the discrete time processes corresponding to feasibly observable data. Section IV derives maximum likelihood estimators for the more general case of mean-reverting processes, allowing for missing observations, and describes how to obtain asymptotic standard errors. Section V performs simulation testing, both to verify large sample and to explore small sample properties of the estimators. In section VI, the estimation procedure is applied to a dataset of stock prices and time varying features of stock return volatility and correlation are investigated. The gain from using incomplete data estimators is also illustrated in this section. The final section provides concluding remarks. Estimators for the special cases of complete data and zero drift model are presented in appendices.

I. On the applications of equicorrelation modeling

Modeling of average sample correlation (mean of off-diagonal elements of the correlation matrix), or equicorrelation (which results from assuming the same pairwise correlation between all the assets in a portfolio), has recently received increasing attention in financial economics. Though a thorough survey of models and applications of equicorrelation is beyond the scope of this paper, we present the major areas of growing interest.

1. Portfolio selection

The covariance matrix of stock returns along with the vector of expected stock excess returns are the main inputs in portfolio selection models. Sample covariance matrix is the standard statistical estimator for the covariance matrix, but, as well documented by Jobson and Korkie (1980), it suffers from severe sampling errors. This problem is amplified when the length of the time series is small compared to the size of the portfolio. This is a typical problem in many financial applications as the length of the time series available for estimation is limited to the shortest-lived stock in the portfolio. Moreover, the sample covariance matrix typically includes a large number of correlation parameters which are almost impossible to interpret in asset allocation applications.

One way to cope with the shortcomings of the sample covariance matrix is to impose some structure such as equicorrelation. Equicorrelation is obtained by assuming the same degree of correlation between all the assets in the portfolio. This leaves fewer parameters to estimate and controls for the impact of extreme sample covariance parameters. The first application of equicorrelation in portfolio management goes back to Elton and Gruber (1973). They showed that assuming the same pairwise correlation among all assets results in superior asset allocation and lower estimation errors over a wide range of alternative correlation structures. Equicorrelation as the only scalar which summarizes the degree of comovements in the stock prices is also quite simple to interpret by portfolio managers.

Some authors, like Ledoit and Wolf (2003, 2004) and Disatnik and Benninga (2007), however, argue that the estimation of the structured covariance matrix introduces specification errors, so propose a combination of structured and sample covariance. Ledoit and Wolf (2003, 2004) suggest shrinking extreme values of the sample covariance matrix toward the center. The shrinkage covariance matrix, Σ_{shrink}, is defined as

$$\Sigma_{shrink} = \alpha\Sigma_{struct} + (1-\alpha)\Sigma_{sample}$$

where Σ_{struct} is the estimated covariance matrix from a highly structured (like equicorrelation) model, Σ_{sample} is the sample covariance matrix and α is the shrinkage parameter. Ledoit and Wolf show that if implemented properly, the shrinkage covariance matrix results in superior portfolio allocations.

Other studies simply use an equally weighted average of the sample and other estimate of the covariance matrix (e.g. set $\alpha = 0.5$). Look at Disatnik and Benninga (2007) and references provided there for a review of this method which they call *portfolio of covariance estimators*. Disatnik and Benninga compare the relative performance of the shrinkage method and portfolio of covariance estimators, and conclude there is no real gain from applying the more complicated shrinkage method.

Whether one prefers the shrinkage or the portfolio of covariance estimator, the question remains to what target should the sample covariance be shrunk, or simply what is the best candidate for Σ_{struct}.[1] Our estimates of equicorrelation could serve both as a target value in shrinkage and portfolio of covariance estimators methods and as a direct indicator of comove-

[1] See Disatnik and Benninga (2007) for a review of the target values used by different authors.

ments in asset portfolios.

2. Derivatives valuation

Equicorrelation is also a common assumption in derivatives pricing. In a dispersion trading strategy, for example, which consists of a long (short) option position on an index of stocks and short (long) option positions on all its constituents where all the positions are delta-hedged, the return will solely depend of the correlations between the stocks. The industry standard approach in computing implied correlation is to assume an equicorrelation structure.

Equicorrelation is even more common in the area of credit derivatives. The assessment of joint default probability in a credit portfolio is the natural starting point for valuing credit derivatives and hedging. Gaussian copula is frequently assumed for building the joint default distribution by both practitioners and academics, in the context of both structural and reduced-form credit risk models. The most important parameter in any copula is the correlation structure, which is either calibrated from market prices of credit derivatives (like *Collateralized Debt Obligations* (CDO's)) or estimated from historical data. Calibration techniques usually assume equicorrelation between obligors in a pool of assets or in a single tranche. Engle and Kelly (2009) review equicorrelation modeling in derivatives pricing.

II. 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

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	$(\rho_i\rho_j)^{1/2}$	$2n$

Notes: the table presents different correlation structres and number of paramters to be estimated under each of them.

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.[2] This structure includes the cases of zero, constant (as a 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 1 (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

[2]The process extends to situations of $-1/(n-1) < \rho < 0$ by defining $dz_0 \equiv \sum_1^n dz_i$ and changing the random term in (2) to $\sigma((1-\rho)^{1/2}dz_i - ((1-\rho)^{1/2} - (1+n\rho-\rho)^{1/2})/ndz_0)$. For more negative ρ, the covariance matrix is not positive-definite.

the estimation group to permit application of results to another group. Thus only the first is considered here.

III. 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

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

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

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

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 $Ve^D V^{-1}$, where V is a matrix whose columns are the eigenvectors of A, and e^D is a diagonal matrix with elements e^{λ_i}, where the λ_i are the corresponding eigenvalues of A. Note that the eigenvectors of hA are the same as for A but with corresponding eigenvalues of $h\lambda_i$. For the process of equation (2), these components are

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

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

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

Substituting these relations into equation (4) and rearranging,

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

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 let n_j denote the number of state variables observed at both (equally spaced) time t_j and time t_{j+1}, $j = 1 \ldots T+1$. It indicates the situation with partial observation series for some or all of the state variables which is quite likely to happen in the financial applications. 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 e_j denote a column vector of ones with appropriate length n_j, and define the n_j-vector

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

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

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

where from equation (7)

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

Maximizing L is equivalent to maximizing Λ, defined as

$$\Lambda \;\equiv\; 2\ln L \;=\; -\Sigma n_j \ln 2\pi - \Sigma \ln|\tilde{\Omega}_j| - \sum_{j=1}^{T} y_j'\tilde{\Omega}_j^{-1}y_j \tag{12}$$

We are almost there. One may verify that that $\tilde{\Omega}_j$ satisfies the following:[3]

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

Substituting into Λ gives

$$\begin{aligned}\Lambda \;=\; & -\Sigma n_j \ln 2\pi - \Sigma(n_j-1)\ln(1-\rho) - \Sigma n_j \ln s - \Sigma_j \ln(1+n_j\rho-\rho) \\ & -\frac{1}{(1-\rho)s}\Sigma y_j'y_j + \frac{\rho}{(1-\rho)s}\Sigma\frac{y_j'\mathrm{e}\mathrm{e}'y_j}{1+n_j\rho-\rho}\end{aligned} \tag{14}$$

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

Summations are understood to be over $j = 1, \ldots, T$. 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. Let $N \equiv \Sigma n_j$, $M \equiv \Sigma g_j n_j^2$ and $g_j \equiv 1/(1 + n_j\rho - \rho)$ to simplify notation.

$$\begin{aligned}\Lambda \;=\;& -N\ln 2\pi - N\ln s - (N-T)\ln(1-\rho) + \Sigma \ln g_j \\ &- \frac{b^2N + \Sigma\tilde{x}_j'\tilde{x}_j - 2a\Sigma x_j'\tilde{x}_j + a^2\Sigma x_j'x_j + 2ab\Sigma x_j'\mathrm{e} - 2b\Sigma\tilde{x}_j'\mathrm{e}}{s(1-\rho)} \\ &+ \frac{\rho(b^2M + \Sigma g_j\tilde{x}_j'\mathrm{ee}'\tilde{x}_j - 2a\Sigma g_j x_j'\mathrm{ee}'\tilde{x}_j + a^2\Sigma g_j x_j'\mathrm{ee}'x_j + 2ab\Sigma g_j n_j x_j'\mathrm{e} - 2b\Sigma g_j n_j\tilde{x}_j'\mathrm{e})}{s(1-\rho)}\end{aligned} \tag{15}$$

IV. Maximum likelihood estimators: mean reversion model

For the general mean-reverting process case, the likelihood function is as in (15). Unlike the zero drift model presented in Appendix B, the ML estimators for the mean-reverting process model 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

$$a = \frac{(\Sigma x_j'\mathrm{e} - \rho\Sigma g_j n_j x_j'\mathrm{e})(\Sigma\tilde{x}_j'\mathrm{e} - \rho\Sigma g_j n_j\tilde{x}_j'\mathrm{e}) + (\Sigma x_j'\tilde{x}_j' - \rho\Sigma g_j x_j'\mathrm{ee}'\tilde{x}_j)(\rho M - N)}{(\Sigma x_j'\mathrm{e} - \rho\Sigma g_j n_j x_j'\mathrm{e})(\Sigma x_j'\mathrm{e} - \rho\Sigma g_j n_j x_j'\mathrm{e}) + (\Sigma x_j'x_j - \rho\Sigma g_j x_j'\mathrm{ee}'x_j)(\rho M - N)} \tag{16}$$

$$b = \frac{(\rho\Sigma g_j n_j\tilde{x}_j'\mathrm{e} - \Sigma\tilde{x}_j'\mathrm{e}) - a(\rho\Sigma g_j n_j x_j'\mathrm{e} - \Sigma x_j'\mathrm{e})}{\rho M - N} \tag{17}$$

$$s = \frac{\Sigma y_j'y_j}{N(1-\rho)} - \frac{\rho\Sigma g_j y_j'\mathrm{ee}'y_j}{N(1-\rho)} \tag{18}$$

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 (16) to (18) to get $\hat{a}, \hat{b}, \hat{s}$.[4]

[4]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

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[5]

$$\hat{\epsilon}_j \;=\; \frac{1}{n_j}\sum_{i=1}^{n_j}(\tilde{x}_{ij} - \hat{a}x_{ij} - \hat{b}) \tag{19}$$

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.[6]

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:[7]

$$T^{1/2}(\hat{\gamma}-\gamma) \;\xrightarrow{d}\; \mathrm{N}(\mathbf{0}, \lim(I/T)^{-1}) \tag{20}$$

in which I denotes the information matrix

$$I = -E(\frac{\partial^2 \ln L}{\partial\gamma\partial\gamma'}) \tag{21}$$

L is the sample size T likelihood function evaluated at the *true* γ, and the limit in (20) 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, Hall and Hausmann as given in Judge (1985, p.180, eq.5.6.8). Their estimator for $\lim(I(\gamma)/T)$ is

$$\frac{1}{T}\left[\sum_{t=1}^{T}\left(\frac{\partial \ln L_t}{\partial\gamma}\right)\left(\frac{\partial \ln L_t}{\partial\gamma}\right)'\right]_{\gamma=\hat{\gamma}} \tag{22}$$

---

expressions involving sixth moments of the data. Numerical maximization then seemed the more expedient route.

[5]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 $\kappa > 0$.

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

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

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	2.4628e+03
maximum	1.1221	5.2943	1.0552	0.3273	2.8386e+03
sample st. dev.	0.0267	0.0966	0.0193	0.0291	6.5957e+01
BHHH st. dev.	0.0277	0.1039	0.0203	0.0288	

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

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.[8] Standard errors for the parameters are square roots of the corresponding diagonal elements of the inverse of this matrix.

V. 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. 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.

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

[8] 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.

agrees closely with the true value used to create the data. Furthermore, the Berndt-Hall-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 occurring 'trend' 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 BHHH 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 BHHH 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.

VI. Application to stock prices

This final section applies the estimation procedure to time series of stock prices of North America oil and gas companies listed on NYSE. List of the companies is obtained from NYSE symbols file[9] and the monthly stock prices from COMPUSTAT dataset. Our sample only includes companies with recorded stock prices for the full period of December, 2002, to March, 2011. We thus start with times series of 100 monthly observations for 99 firms. The companies are further grouped into 8 sub-industries, according to NYSE symbols file: oil and gas-contract drilling, exploration (CDE), oil and gas-crude production (CPR), oil and gas-integrated domestic refiners (IDR), oil and gas-integrated international refiners (IIR), oil and

[9]The file is available at: http://www.nyxdata.com/Data-Products/NYSE-Group-Symbols-Package.

Table 3: Simulation with 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
BHHH 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
BHHH 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
BHHH 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
BHHH 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
BHHH 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
BHHH st. dev.	5.7953	14.6454	8.6073	6.7456	

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

Table 4: Simulation with 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
BHHH 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
BHHH 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
BHHH 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
BHHH 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
BHHH 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
BHHH st. dev.	0.2059	0.2304	0.0847	0.1474	

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

gas-services and equipment (S&E), gas services-distribution & integrated natural gas (DIN), gas services-natural gas transmission companies (NGT), and other gas services(OGS). Descriptive statistics of log-returns are summarized in Table 5. Average sample correlation is the average of off-diagonal elements of the sample correlation matrix. As can be seen, average sample mean and correlation of OGS and DIN, and volatility of DIN are noticeably smaller than those of the other sub-industries.

The goal is to estimate the model parameters, with σ and ρ being the focus of interest. As the estimates of σ and ρ turn out to be fairly insensitive to the specification of the drift term, we just report results based on the constant drift model. Table 6 presents maximum likelihood estimates of the process parameters under the assumption of constant drift. Estimates based on zero drift is reported for the whole sample to illustrate the insensitivity of the σ, ρ estimates to the drift specification. As expected, estimates of μ match the sample means reported in Table 5. Similar to what we observed in Table 5, the estimated equicorrelation for DIN and OGS differ significantly from the other sub-industries. Estimated ρ for NGT also turns out to be notably smaller than the average sample correlation and estimated ρ for the first 5 sub-sectors. Therefore, estimation for the whole sample is performed once excluding the three gas services sub-sectors (i.e. DIN, NGT, OGS). CPR and IDR are the most correlated while DIN and OGS are the least correlated groups. Equicorrelations of the first five sub-industries are higher than those of the other three, ranging from 0.27 to 0.37.

Estimated equicorrelations generally differ from the average sample correlations presented in Table 5. This is important, since the estimates used would alter choices in portfolio selection and risk management. Note also that the average sample correlations show lower variability across sub-industries. This suggests that average sample correlations may not be able to capture sector-specific variability of equicorrelation. Further research is needed to shed light on the implications of replacing average sample correlations by the suggested ML estimates for portfolio allocation.

To investigate the variability of equicorrelation, rolling estimates are obtained for 24-month windows with one month steps. Results are reported in Figure VI.. A significant increase in stock return correlation could be seen in 2008, coinciding with the peak of the global financial crisis. This happens somewhat earlier in gas services sub-sectors. The rise in correlation is

Table 5: Stock log-return sample properties

	No. of diffusions	Mean	Volatility	Avg. Corr.	Min	Max
CDE	25	0.1099	0.5163	0.3731	-2.1735	3.0884
CPR	27	0.1097	0.4231	0.3722	-1.4380	0.6732
IDR	5	0.1377	0.4938	0.3775	-0.6546	0.6660
IIR	10	0.1017	0.3763	0.3339	-1.0432	0.3627
S&E	26	0.1092	0.4345	0.3366	-1.3702	1.1793
DIN	20	0.0659	0.2922	0.1858	-1.0172	1.3692
NGT	9	0.0885	0.3979	0.3247	-0.9760	0.8523
OGS	6	-0.0048	0.5143	0.1754	-1.3780	0.5959
excluding gas sub-sectors*	93	0.1102	0.4501	0.3389	-2.1735	3.0884
whole sample	128	0.0964	0.4248	0.2850	-2.1735	3.0884

Notes: the sample includes North American firms listed on NYSE. List of companies is obtained from NYSE Symbols file, and only companies with recorded stock prices over the full period of Dec. 2002 to May. 2011 are included. Sample mean and volatility are annualized.

* This sample excludes DIN, NGT, OGS sub-sectors.

Table 6: Estimation results for stock prices

	μ	σ	ρ	Λ^*
CDE	0.1099	0.5609	0.2705	2.5387e+03
	(0.116)	(0.007)	(0.021)	
CPR	0.1097	0.4441	0.3539	4.2905e+03
	(0.096)	(0.011)	(0.035)	
IDR	0.1377	0.4972	0.3733	6.1173e+02
	(0.130)	(0.009)	(0.038)	
IIR	0.1017	0.3887	0.2895	1.6993e+03
	(0.083)	(0.004)	(0.022)	
S&E	0.1092	0.4691	0.2940	3.6394e+03
	(0.102)	(0.008)	(0.029)	
DIN	0.0659	0.3265	0.1182	3.8544e+03
	(0.047)	(0.002)	(0.013)	
NGT	0.0885	0.4475	0.2242	1.2178e+03
	(0.090)	(0.006)	(0.034)	
OGS	-0.0048	0.5447	0.1267	5.3052e+02
	(0.100)	(0.004)	(0.036)	
whole sample excluding gas	0.1102	0.4826	0.2901	1.2959e+04
sub-sectors (constant-drift)	(0.102)	(0.009)	(0.026)	
whole sample (constant-drift)	0.0964	0.4625	0.2430	1.8228e+04
	(0.096)	(0.007)	(0.025)	
whole sample (zero-drift)		0.4633	0.2457	1.8226e+04
		(0.007)	(0.024)	

Notes: this table presents the estimation results of the constant drift model for different sub-sectors. Volatility and equicorrelation estimates turn out to be insensitive to the drift specification (this is observable from the estimation of zero drift model for the whole sample.) Numbers in the parentheses are BHHH standard errors of parameters. Whole sample excluding gas sub-sectors is formed by removing DIN, NGT and OGS from to whole sample to resemble a more homogeneous portfolio.

most obvious in CDE, IRR, S&E , DIN and the two all-industry samples, while is quite weak in the other sub-industries (with IDR being the weakest). In the first group of sub-industries, this means that a higher proportion of stock return variations is explained by the common factor, or simply indicates that the financial crisis brought higher systematic risk to the stock market. In all of the graphs (except IDR and IRR), the correlation tends to decrease moving toward the end of the sample period. This rising/falling pattern is weaker for CPR, NGT and OGS, but quite obvious for CDE, S&E, DIN and the two whole samples. More specifically, in the first sub-industry group the two times standard error band is wide enough to include most of the variation in correlation over time.

Application to incomplete data

We now turn to applying the estimators to incomplete datasets, created by randomly removing some data points from the sample of oil and gas companies. Without the incomplete data estimators presented in this paper, there are two alternatives for handling incomplete data: either the months or the firms with some missing data points must be removed to convert the dataset with partial observations to a smaller sample with no missing data. However, in this process some existing data is thrown way. For a less homogeneous portfolio (including assets with wide range of volatility and pair-wise correlation), more valuable information is missed by removing the firms with missing data. Whereas if the missing periods are less overlapping across firms, removing the periods with missing observations becomes more costly. Our partial data estimators permit all the available information to be utilized in the estimation.

To examine the relative performance of these three alternative methods, the following experiment is performed. First, 50 firms are randomly picked from the whole sample of oil and gas companies (excluding DIN, NGT, OGS)[10] and 50 consecutive observations of all the selected firms are removed, with the starting time interval being randomly picked[11]. The resulting sample with partial data series is then used to estimate the parameters via the

[10] These three sub-sectors are excluded to work with a more homogeneous sample.

[11] We don't let the starting point of missing observations vary across companies since it may result in a null dataset when, following the first alternative method, all the months with missing data are removed.

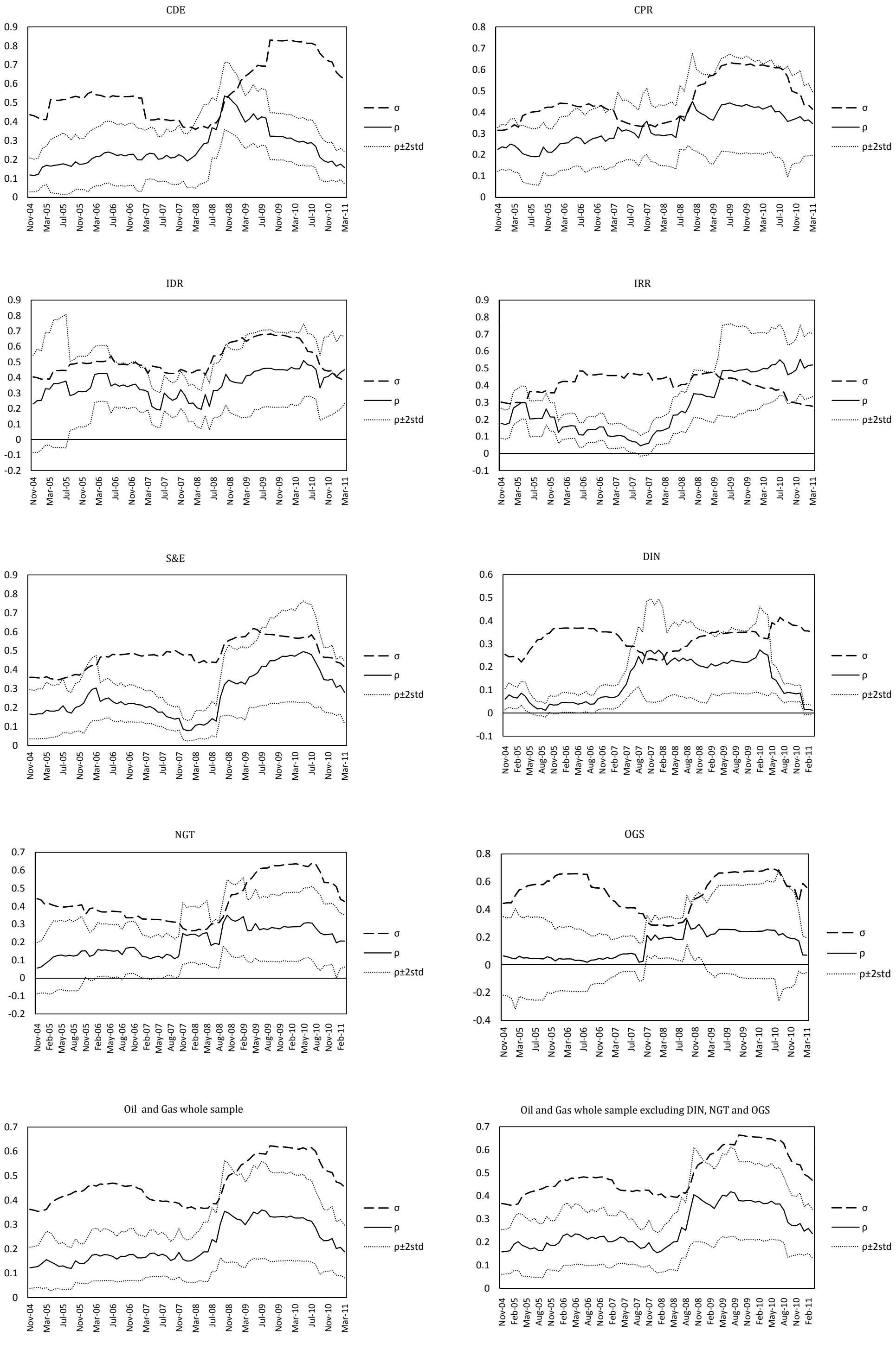

Figure 1: Rolling estimation of equicorrelation and volatility

Notes: parameter estimates are obtained from rolling estimation of the constant drift model over 24-month windows with one-month steps. Thus, estimated parameters for each month (say, November 04) are based on 24-month time series ending in that month (December 2002 to November 2004).

incomplete data estimators and the two alternative methods. This experiment is repeated 100 times and the averages of *mean square errors* (MSE) of the estimated σ and ρ for each method are reported in the upper part of Table 7. MSEs are calculated around the parameter estimates obtained from the complete sample (reported in the lower part of Table 6). As can be seen, MSE is relatively much larger if someone removes all the months with missing observations. It is also smaller if one uses the incomplete data estimators than if one removes completely the firms with missing data.

Given that sample is fairly homogeneous, the small MSE of the third method is somewhat expected. The lower part of Table 7 reports the result of a similar experiment for a less homogeneous portfolio consisting of only the PRO and DIN sub-industries. In this second experiment, 50 consecutive months of 23 firms are removed each time in a similar manner to the first experiment. As can be seen, the MSE for the third method increases substantially and becomes about 3 times as big as the MSE of the incomplete data estimators. In summary, the incomplete data estimators generally produce closer estimates to the values obtained if no data was missing. The gain from using these estimators increases as the number of missing months increases and as the portfolio becomes less homogeneous. In the extreme cases of short and frequent periods of missing observations in majority of the state variables, incomplete data estimators would be the only feasible way to estimate the parameters of the correlated diffusions.

Dealing with outliers

Outliers can cause substantial bias in parameter estimation and serious problems in time series analysis. The incomplete data estimators can be used to deal with this problem. A formal test of the incomplete data estimators performance in handling outliers involves using legitimate methods of identifying them and comparing the results to the available solutions. For the sake of illustration, we identify two data points which could be considered outliers by looking into the sample of stock prices: minimum and maximum values of the sample, which both lie in CDE sub-industry. Table 8 presents the results after removing these two data points; this removes 4 data points from the n-vector $y_j (j = 1, \ldots, T)$. As can be seen, estimates of the parameters are noticeably different when the outliers are removed.

Table 7: MSE of estimates for incomplete datasets of stock prices

Sample	Method of handling incomplete data	σ	ρ
Whole sample	I. Incomplete data estimators	.0006775	.0004646
	II. Removing months with missing data	.0277	.0140
	III. Removing firms with missing data	.0007854	.0007154
CPR and DIN	I. Incomplete data estimators	.0002835	.0002939
	II. Removing months with missing data	.0134	.0157
	III. Removing firms with missing data	.0005973	.0008130

Notes: the table reports mean square errors of three alternative methods of handling incomplete datasets. The upper part of the table presents MSE of estimates from the whole sample excluding gas sub-sectors (DIN, NGT, OGS), and the lower section reports MSEs for a less homogeneous portfolio consisting of CPR and DIN. To create incomplete datasets from the whole sample, 50 firms are randomly picked and 50 consecutive observations of all the selected firms are removed, with the starting time interval being randomly determined. The same procedure with 23 firms is followed to create incomplete datasets from CPR and DIN sample. This experiment is repeated 100 times and MSE is computed around the estimated σ, ρ from complete samples.

Table 8: Estimation results after removing the outliers

	μ	σ	ρ	Λ^*
CDE (including outliers)	0.1099	0.5609	0.2705	2.5387e+03
CDE (removing outliers)	0.1058	0.4955	0.3255	3.3170e+03
	(0.109)	(0.008)	(0.026)	
whole sample (including outliers)	0.0964	0.4625	0.2430	1.8228e+04
whole sample (removing outliers)	0.0956	0.4477	0.2556	1.9254e+04
	(0.101)	(0.009)	(0.028)	

Notes: the table contains estimation results after removing the outliers. For the ease of comparison, estimates with outliers are also reported. The sample min and max stock returns are treated as outliers. Numbers in the parentheses are BHHH standard errors.

VII. Conclusion

This paper suggests 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 feasibility of maximum likelihood method.

The estimation procedure is applicable to systems of any size, with missing observations of any pattern. This provides a solution to the situations where one needs to deal with large asset portfolios while value of some or all of them are not observed at some times during the life of the portfolio. The procedure was applied to the stock price returns of the North America oil and gas companies traded on NYSE and time varying features of equicorrelation was investigated. The performance of the estimation procedure on partially observed time series was also examined. The incomplete data estimators outperform (by generating smaller MSE) the estimates based on simply removing rows or columns with missing observations from the data matrix. It is left to future studies to investigate the implications of using the estimators suggested in this paper for portfolio selection, risk management and derivatives pricing.

References

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

[2] Disatnik, David J., and Simon Benninga, "Shrinking the Covariance Matrix: Simpler is Better," *Journal of Portfolio Management*, 33 (2007), 55-63.

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

[4] Elton, Edwin J., and Martin J. Gurber, "Estimating the Dependence Structure of Share Prices-Implications for Portfolio Selection," *Journal of Finance*, 28 (1973), 1203-1232.

[5] Engle, Robert, and Kelly Bryan, "Dynamic Equicorrelation," NYC Stern School of Business Working Paper, 2009.

[6] Engle, Robert, " Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models," *Journal of Business and Economic Statistics*, 20 (2002), 339-350.

[7] 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.

[8] Jobson, J. Dave, and Bob M. Korkie, "Estimation for Markowitz Efficient Portfolios," *Journal of the American Statistical Association*, 75 (1980), 544-554.

[9] Ledoit, Olivier, and Michael Wolf, "Improved Estimation of the Covariance Matrix of Stock Returns with an Application to portfolio selection," *Journal of Empirical Finance*, 10 (2003), 603-621.

[10] Ledoit, Olivier, and Michael Wolf, "Honey, I Shrunk the Sample Covariance Matrix," *Journal of Portfolio Management*, 30 (2004), 110-119.

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

[12] Wymer, C.R., "Econometric Estimation of Stochastic Differential Equation Systems," *Econometrica*, 40(3) (1972), 565-578.

Appendix

A. Complete data series

This appendix presents the likelihood function and the ML estimators for the case of no missing data. Suppose we have observations on all n state variables at equally spaced times t_j $(j = 1 \dots T)$, with $y_j = x_{j+1} - ax_j - b\mathrm{e}$ being the n-vector of random parts of the changes as in equation (9). The overall likelihood function is as in equation (10) with n_j becoming n and recognizing that the covariance matrix $\tilde{\Omega}$ is of size n for all j. Making this adjustment results in the revised two-times log-likelihood function corresponding to equation (14):

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

Substituting for y_j and collecting terms allows Λ to be written in terms of the parameters and data moments as

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

Summations are understood to be over $j = 1, \dots, T$ and the e is n-vectors of 1's.

For given ρ, this may be maximized with respect to a, b, s by setting first partial derivatives equal to 0 and solving for their values

$$a = \frac{nT(1+n\rho-\rho)\Sigma x_j'\tilde{x}_j - \rho nT\Sigma x_j'\mathrm{ee}'\tilde{x}_j - (1-\rho)\Sigma x_j'\mathrm{e}\Sigma\tilde{x}_j'\mathrm{e}}{nT(1+n\rho-\rho)\Sigma x_j' x_j - \rho nT\Sigma x_j'\mathrm{ee}'x_j - (1-\rho)\Sigma x_j'\mathrm{e}\Sigma x_j'\mathrm{e}} \tag{A.3}$$

$$b = \frac{\Sigma\tilde{x}_j'\mathrm{e} - a\Sigma x_j'\mathrm{e}}{nT} \tag{A.4}$$

$$s = \frac{\Sigma y_j' y_j}{nT(1-\rho)} - \frac{\rho\Sigma y_j'\mathrm{ee}'y_j}{(1-\rho)(1+n\rho-\rho)} \tag{A.5}$$

The concentrated likelihood function $\Lambda^*(\rho)$ is obtained by substituting these values into (A.3). 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.

B. Zero drift model

This appendix specializes to the case of zero drift by setting $a = 1$ and $b = 0$ for the case of no missing data. This could be readily extended to the case of partial data series. The likelihood function reduces to

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

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

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

$$\Lambda^* = -T\left((\ln 2\pi + \ln\hat{s} + 1)n + (n-1)\ln(1-\hat{\rho}) + \ln(1+n\hat{\rho}-\hat{\rho})\right) \qquad \text{(B.3)}$$

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

$$\hat{\sigma} = (\hat{s}/h)^{1/2} \qquad \text{(B.4)}$$

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

$$\hat{\epsilon}_j = \frac{1}{n}\sum_{i=1}^{n}(\tilde{x}_{ij} - x_{ij}) \qquad \text{(B.5)}$$

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.