Estimating Volatility in the Merton Model
The KMV Estimate is not Maximum Likelihood

Christoffersen, Benjamin; Lando, David; Nielsen, Søren Feodor

Document Version
Final published version

Published in:
Mathematical Finance

DOI:
10.1111/mafi.12362

Publication date:
2022

Received: 11 June 2021 | Accepted: 13 July 2022

DOI: 10.1111/mafi.12362

ORIGINAL ARTICLE

MATHEMATICAL FINANCE An International Journal of Mathematics, Statistics and Financial Economics

WILEY

Estimating volatility in the Merton model: The KMV estimate is not maximum likelihood

Benjamin Christoffersen[1,2] | **David Lando**[3] | **Søren Feodor Nielsen**[3]

[1]KTH Royal Institute of Technology, Stockholm, Sweden

[2]Karolinska Institutet, Stockholm, Sweden

[3]Department of Finance, Copenhagen Business School, Frederiksberg, Denmark

Correspondence
David Lando, Department of Finance, Copenhagen Business School, Solbjerg Plads 3, DK-2000 Frederiksberg, Denmark.
Email: dl.fi@cbs.dk

Funding information
Danmarks Grundforskningsfond, Grant/Award Number: DNRF102

Abstract

We compare two methods for estimating the asset volatility in the Merton model using observed equity prices: maximum likelihood and an iterative method commonly referred to as the KMV method. The two methods often yield extremely similar estimates, which has led to the conjecture that the two methods are equivalent. We show that this is not true and we provide a necessary and sufficient condition that the inverse of the equity pricing function would have to satisfy for the two methods to be equivalent. Moreover, we show numerically that this condition is very close to being true for in-the-money options.

KEYWORDS

distance-to-default, EM-algorithm, KMV method, Merton model

1 | INTRODUCTION

The classical (Merton, 1974) model for pricing corporate bonds views debt and equity issued by a firm as derivative securities on the market value of the firm's assets. This approach yields a powerful and internally consistent approach to pricing the different elements of a firm's capital structure. It also provides a key statistic—distance-to-default—that has been shown to be a powerful predictor of firm defaults, see for example, Crosbie and Bohn (2003) and Bharath and Shumway (2008). In contrast to standard derivative pricing models, the underlying asset, that is, firm asset value, cannot be directly observed and has to be estimated along with the parameters of the model. Since equity is often the most liquidly traded security in a firm's capital structure, it is common to base the estimation on observations of the equity price process.

Two methods have been proposed for this estimation: a maximum-likelihood estimation (MLE) method proposed in Duan (1994), and an iterative approach reportedly used historically by KMV (now a part of Moody's) and described in Vassalou and Xing (2004). We will refer to the iterative method when applied in the context of a Merton-type model as the KMV method. The KMV method seems to always converge rapidly to values that are often extremely close to those obtained by the MLE approach. This has led to the conjecture that the two methods are in fact equivalent and therefore should produce the same estimates of the parameters, see for example, Duan et al. (2004) who argue that the KMV algorithm may be viewed as an instance of the EM-algorithm.

We show in this paper that the estimates produced by the two methods are not the same. The reason is that the first-order conditions, which must be satisfied at a fixed point of the KMV algorithm are not the same as those for the MLE. The difference in first-order conditions vanishes if the function that inverts the observed values of equity to asset values factors into a product of a function that depends on volatility and a function that depends on the observed equity value. We show that this multiplicative separability condition is necessary and sufficient for the KMV method and the MLE to produce the same estimates.

The separability does not hold for the Black–Scholes inverse function, but we show numerically that it is extremely close to being true for in-the-money options, that is, when firm asset value is large compared to the face value of the firm's debt. The factorization is not quite as close to holding true for out-of-the-money options. This explains why in our simulations we are better able to detect differences between the estimates of the two methods when the equity call option is deep out-of-the-money.

Consistent with our theoretical findings, we verify numerically that when we start the KMV algorithm at the MLE estimates, the algorithm drifts away towards its fixed point, which is a point of lower likelihood. This also indicates that the KMV method cannot be viewed as an implementation of the EM algorithm. If the method were an EM-algorithm, the likelihood would have to increase (or at least not decrease) in each step. But the fact that the KMV method moves away from the MLE estimates implies that it decreases the likelihood. We explain why there is no guarantee in the KMV approach that the likelihood increases in each step, as would be required for it to be an EM-algorithm.

We also present a simple counterexample, which shows that one cannot in general expect an iterative approach such as the KMV method to even converge. Therefore, it is by no means obvious that the KMV method can be expected to have universal applicability.

2 | ESTIMATION IN THE MERTON MODEL

Our study is motivated by the Merton model, and we, therefore, start by briefly reviewing the two methods in the context of this model.

2.1 | The Merton model

The starting point in Merton (1974) is a model for the market value V of a firm's assets, which is assumed to follow a geometric Brownian motion, that is,

$$dV_t = \mu V_t dt + \sigma V_t dW_t,$$

where μ and σ are the drift and volatility of the asset value, respectively, and W_t is a standard Brownian motion. Equity and bonds issued by the firm are viewed as derivative securities with V as underlying security. In the simplest version in which the firm has issued zero coupon bonds with face value D and maturity at time T, the pay-off to equity at the maturity date is given as

$$S_T = \max(V_T - D, 0),$$

which is equivalent to the pay-off of a European call option on V with strike price D. If we assume that there is a riskless security in the economy with a constant riskless rate equal to r, then we can use the Black–Scholes formula to price equity by

$$S_t = C(V_t, D, \sigma, r, T - t), \tag{1}$$

where $C(V_t, D, \sigma, r, T - t)$ denotes the Black–Scholes price of a European call option with strike price D and time to maturity $T - t$, that is,

$$C(V_t, D, \sigma, r, T - t) = V_t \Phi(d_1) - D \exp(-r(T - t))\Phi(d_2),$$

with

$$\begin{aligned} d_1 &= \frac{\log(V_t/D) + \left(r + \sigma^2/2\right)(T - t)}{\sigma\sqrt{T - t}} \\ d_2 &= d_1 - \sigma\sqrt{T - t}, \end{aligned} \tag{2}$$

where Φ denotes the cumulative distribution function of the standard normal distribution.

We stress that the underlying security—the firm's asset value—is not observable and must be inferred from data on equity prices. Our focus is the two methods for doing this which we now review.

2.2 | The KMV method

To understand the KMV method, we first consider the expression for the likelihood function that we could compute if we had observed the underlying asset values: The dates of observation are $0 = t_0 < t_1 < \cdots < t_n \leq T$ and for simplicity we will assume that all time increments are equal and write $\Delta t = t_i - t_{i-1} = T/n$. We will also use the shorthand V_i instead of V_{t_i} (and similarly for S and W) in all expressions for the likelihood function. With this notation, we can write

$$V_i = V_{i-1}\frac{V_i}{V_{i-1}} = V_{i-1}\exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t + \sigma(W_i - W_{i-1})\right).$$

Since the asset value process is not stationary, we obtain the likelihood function by conditioning on $V_0 = v_0$ which leads to

$$L_V(\mu,\sigma \mid v_0,\dots,v_n) = \prod_{i=1}^{n} \phi(v_i \mid v_{i-1},\mu,\sigma) \tag{3}$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2\Delta t}} \exp\left[-\frac{\left(\log v_i - \log v_{i-1} - \left(\mu - \sigma^2/2\right)\Delta t\right)^2}{2\sigma^2\Delta t}\right]\frac{1}{v_i},$$

where $\phi(*)$ denotes the density of $V_i \mid V_{i-1}$ for $i = 1,\dots,n$.

In practice, we cannot observe the asset values, but if we knew the volatility of the firm's asset, we could invert the set of observed equity prices into corresponding asset values, and then use this likelihood function to estimate the parameters of the firm value process. The KMV method starts by guessing a value of the volatility, $\sigma^{(0)}$. Using this guess, we obtain a time series of "implied asset values" by inverting Equation (1), which we can then plug into the likelihood function in Equation (3), which we maximize to obtain an estimate of the parameters $(\mu^{(1)},\sigma^{(1)})$. The new estimate, $\sigma^{(1)}$, is then used again to invert the observed equity prices into a time series of asset values, a new maximization of the likelihood is performed—and so forth. In practice, this procedure seems to converge very quickly and the estimates of μ and σ are very often close to the estimates obtained by the MLE method, which we consider next.

2.3 | Likelihood based on equity values

Following Duan (1994), we use a change of variables to derive the likelihood based on observation of equity values as the transformation from S_t to V_t is strictly monotonic and differentiable. Let S_t denote the equity value of a firm at time t, such that $S_t = C(V_t,\sigma)$ where $C(\cdot)$ denotes the Black–Scholes formula (1) with known parameters (strike price, time to maturity, and interest rate) suppressed for notational simplicity. The likelihood function based on equity values (and conditional on $S_0 = s_0$) then becomes

$$L_S(\mu,\sigma \mid s_0,\dots,s_n) = \prod_{i=1}^{n} \phi(v_i(\sigma) \mid v_{i-1}(\sigma),\mu,\sigma)\left|\frac{1}{C'(v_i(\sigma),\sigma)}\right|, \tag{4}$$

where $v_i(\sigma)$, the value of the firm's assets, is obtained by inverting the Black–Scholes formula (1) using the observed equity price, s_i, and volatility σ. With a fixed maturity date T, the time to maturity is not the same for the different observations, but the index i on $v_i(\sigma)$ refers to the observation date t_i and hence implies a time to maturity of $T - t_i$. C' is the derivative of the Black–Scholes function w.r.t. the value of the underlying (i.e., the option delta). Maximizing Equation (4) w.r.t. σ and μ yields the MLE estimates.

3 | THE ITERATIVE METHOD NEED NOT CONVERGE

While the KMV method is intuitively appealing and shows strong convergence properties in practice, it is by no means obvious that the iterative approach always works. In this section, we provide a simple example of how such an iterative approach may fail. To make a formal analysis feasible, we work with an example in which, in contrast to the Merton model, the inversion from the observed data (equity) to the underlying variable (asset value) can be done analytically.

Let $V_1, \dots, V_n$ be iid exponentially distributed with mean σ. The likelihood is

$$L_V(\sigma | V_1, \dots, V_n) = \frac{1}{\sigma^n} \exp(-n\bar{V}/\sigma), \tag{5}$$

with $\bar{V}$ denoting the average of the V_i's. This likelihood is maximized for $\sigma = \bar{V}$. Suppose that the V_is are unobservable, and that we instead observe $S_i = \sigma V_i$, $i = 1, \dots, n$. Then S_i is exponentially distributed with mean σ^2 and the likelihood equals

$$L_S(\sigma | S_1, \dots, S_N) = \frac{1}{\sigma^{2n}} \exp(-n\bar{S}/\sigma^2),$$

with $\bar{S}$ denoting the average of the S_i's. The MLE estimate of σ based on observation of $S_1, \dots, S_n$ equals $\hat{\sigma} = \sqrt{\bar{S}}$.

The "KMV-like" algorithm would calculate $\hat{V}_i = S_i/\hat{\sigma}^{(0)}$ using the initial estimate $\hat{\sigma}^{(0)}$, plug these "implied values" into the likelihood (5), and maximize this to obtain $\hat{\sigma}^{(1)} = \bar{S}/\hat{\sigma}^{(0)}$. Repeating this would lead to $\hat{\sigma}^{(2)} = \bar{S}/\hat{\sigma}^{(1)} = \sigma^{(0)}$, that is, an alternating sequence of estimates that would never converge unless started at $\hat{\sigma}^{(0)} = \sqrt{\bar{S}}$. Hence, the algorithm only converges to its fixed point $\sqrt{\bar{S}}$ when started at the fixed point. All other starting points are periodic points of order 2.

4 | KMV FIXED POINT IS NOT MLE

Having seen that an iterative method like the KMV method need not converge, we now turn more specifically to the KMV method. If it converges, it will converge to a fixed point. Our first theorem shows that such a fixed point, if it exists, is not a solution to the likelihood equation. To see this, we let $G(s, \sigma)$ denote the inverse of the pricing function, that is, the function calculating the value of the underlying asset, v. This function need not be the Black–Scholes function, but we do assume that it is strictly positive. Our first theorem shows that G would have to be multiplicatively separable in its two arguments s and σ (that is, factor into a product of some function of s and some function of σ) for the fixed point of the KMV method to be a solution to the likelihood equations. Note that multiplicative separability of the inverse function is the same as additive separability of the logarithm of the inverse function.

Theorem 4.1. *Let s_i denote the equity value observed at date t_i and let $G(s_i, \sigma)$ denote the value of the underlying assets v_i for which*

$$C(v_i, \sigma, K, T - t_i, r) = s_i,$$

where C is a general pricing formula.

If $\log G(s, \sigma)$ is not additively separable, then a fixed point of the KMV algorithm is not a solution in general to the likelihood equation for σ, that is, the first-order conditions for the MLE of σ.

Proof. Using superscripts to denote the iteration number of the KMV method, we consider the computation of the k'th iteration based on the estimate $\sigma^{(k-1)}$. (The drift term $\mu^{(k-1)}$ does not influence the next step).

To ease notation, we reparametrize the model using $m = \mu - \sigma^2/2$. Then, in the k'th iteration, the function we need to maximize with respect to m and σ is given by

$$-\frac{n}{2}\log(2\pi\Delta t) - n\log\sigma + \sum_{i=1}^{n}\left(-\frac{\left(\log\frac{G(s_i,\sigma^{(k-1)})}{G(s_{i-1},\sigma^{(k-1)})} - m\Delta t\right)^2}{2\sigma^2\Delta t} - \log G(s_i,\sigma^{(k-1)})\right).$$

The first-order condition for m implies that

$$m^{(k)} = m(\sigma^{(k-1)}),$$

with

$$m(\sigma) = \frac{1}{n\Delta t}\sum_{i=1}^{n}\log\frac{G(s_i,\sigma)}{G(s_{i-1},\sigma)}. \tag{6}$$

Thus, the first-order condition for σ is

$$\frac{\partial}{\partial\sigma}\left(n\log\sigma + \sum_{i=1}^{n}\frac{\left(\log\frac{G(s_i,\sigma^{(k-1)})}{G(s_{i-1},\sigma^{(k-1)})} - m(\sigma^{(k-1)})\Delta t\right)^2}{2\sigma^2\Delta t}\right) = 0.$$

This leads to the standard result

$$\sigma^{(k)2} = \frac{1}{n\Delta t}\sum_{i=1}^{n}\left(\log\frac{G(s_i,\sigma^{(k-1)})}{G(s_{i-1},\sigma^{(k-1)})} - m(\sigma^{(k-1)})\Delta t\right)^2.$$

Hence, a fixed point σ of the KMV method must satisfy

$$\sigma^2 = \frac{1}{n\Delta t}\sum_{i=1}^{n}\left(\log\frac{G(s_i,\sigma)}{G(s_{i-1},\sigma)} - m(\sigma)\Delta t\right)^2.$$

We will compare this with the first-order condition for the MLE of σ. Dropping constant terms from the log-likelihood, the function we need to maximize to find the MLE estimate is

$$-n\log\sigma + \sum_{i=1}^{n}\left(-\frac{\left(\log\frac{G(s_i,\sigma)}{G(s_{i-1},\sigma)} - m\Delta t\right)^2}{2\sigma^2\Delta t} - u_i(\sigma)\right),$$

with

$$u_i(\sigma) = \log G(s_i,\sigma) + \log|C'(G(s_i,\sigma),\sigma)|.$$

Differentiating with respect to m and solving the resulting first-order condition, we see that m must satisfy Equation (6). Consequently, the profile log-likelihood function for σ is

$$-n\log\sigma + \sum_{i=1}^{n}\left(-\frac{\left(\log\frac{G(s_i,\sigma)}{G(s_{i-1},\sigma)} - m(\sigma)\Delta t\right)^2}{2\sigma^2\Delta t} - u_i(\sigma)\right), \tag{7}$$

which leads to the first-order condition for the MLE of σ:

$$\sigma^2 = \frac{1}{n\Delta t}\sum_{i=1}^{n}\left(\log\frac{G(s_i,\sigma)}{G(s_{i-1},\sigma)} - m(\sigma)\Delta t\right)^2 - \frac{\sigma^3}{n}\left(\frac{1}{2\sigma^2\Delta t}\sum_{i=1}^{n}\frac{\partial}{\partial\sigma}\left(\log\frac{G(s_i,\sigma)}{G(s_{i-1},\sigma)} - m(\sigma)\Delta t\right)^2 + \sum_{i=1}^{n}u_i'(\sigma)\right). \tag{8}$$

Hence, a fixed point, σ, of the KMV method will only solve the likelihood equations, if the second term on the right-hand side of Equation (8) is zero, that is, if

$$\frac{1}{2\sigma^2\Delta t}\sum_{i=1}^{n}\frac{\partial}{\partial\sigma}\left(\log\frac{G(s_i,\sigma)}{G(s_{i-1},\sigma)} - m(\sigma)\Delta t\right)^2 + \sum_{i=1}^{n}u_i'(\sigma) = 0. \tag{9}$$

The second term in this equation does not depend on s_0 and therefore, the first term cannot vary with s_0, if the sum of the two terms has to equal 0. We now show that for the first term to not vary with s_0, $\log G(s,\sigma)$ has to be additively separable.

Before calculating the derivative with respect to σ, we note that

$$\frac{1}{n\Delta t}\sum_{i=1}^{n}\left(\log\frac{G(s_i,\sigma)}{G(s_{i-1},\sigma)} - m(\sigma)\Delta t\right)^2 = \frac{1}{n\Delta t}\sum_{i=1}^{n}\left(\log\frac{G(s_i,\sigma)}{G(s_{i-1},\sigma)}\right)^2 - m(\sigma)^2\Delta t.$$

Also, note that

$$m(\sigma) = \frac{1}{n\Delta t}\sum_{i=1}^{n}\log\frac{G(s_i,\sigma)}{G(s_{i-1},\sigma)} = \frac{1}{n\Delta t}(\log G(s_n,\sigma) - \log G(s_0,\sigma)).$$

Now

$$\begin{aligned}
&\frac{\partial}{\partial\sigma}\left(\frac{1}{n\Delta t}\sum_{i=1}^{n}\left(\log\frac{G(s_i,\sigma)}{G(s_{i-1},\sigma)}\right)^2 - m(\sigma)^2\Delta t\right)\\
&= \frac{2}{n\Delta t}\sum_{i=1}^{n}\left(\frac{G_\sigma'(s_i,\sigma)}{G(s_i,\sigma)} - \frac{G_\sigma'(s_{i-1},\sigma)}{G(s_{i-1},\sigma)}\right)(\log G(s_i,\sigma) - \log G(s_{i-1},\sigma))\\
&\quad - 2m'(\sigma)m(\sigma)\Delta t\\
&= \frac{-2}{n\Delta t}\left(\frac{G_\sigma'(s_1,\sigma)}{G(s_1,\sigma)}\log G(s_0,\sigma) + \frac{G_\sigma'(s_0,\sigma)}{G(s_0,\sigma)}\log G(s_1,\sigma)\right)
\end{aligned}$$

$$+\frac{2}{n\Delta t}\frac{G'_\sigma(s_n,\sigma)}{G(s_n,\sigma)}\log G(s_0,\sigma)+\frac{2}{n\Delta t}\frac{G'_\sigma(s_0,\sigma)}{G(s_0,\sigma)}\log G(s_n,\sigma)$$
$$+\text{ terms not depending on } s_0.$$

Here, $G'_\sigma(s,\sigma)$ denotes the derivative of $G(s,\sigma)$ w.r.t. σ. For this expression to be independent of s_0, $\frac{G'_\sigma(s,\sigma)}{G(s,\sigma)}$ must be independent of s, which can be rephrased as the condition that

$$\frac{\partial}{\partial s}\frac{\partial}{\partial\sigma}\log G(s,\sigma)=0,$$

which amounts to saying that $\log G(s,\sigma)$ is additively separable. □

As we shall see, the inverse Black–Scholes formula is not multiplicatively separable, but in some parts of its domain, it is not far from being the case. This can explain the proximity of the estimators obtained through the two methods we consider. We return to this point below in our section with numerical illustrations.

5 | A CONDITION FOR EQUIVALENCE

We have established that multiplicative separability of the inverse option pricing function is a necessary condition for equivalence of the two methods, but is it also sufficient? Our second theorem provides an answer to this question:

Theorem 5.1. *If the inverse pricing function, G, linking observed equity values to underlying asset values is multiplicative separable, that is, factorizes such that $G(s,\sigma)=f(\sigma)h(s)$ with h a continuously differentiable function, which is strictly positive and monotone, then the estimates obtained from the KMV method are the MLE estimates. Moreover, the KMV method converges in one iteration.*

Proof. With $V_i=G(S_i,\sigma^{(k-1)})=f(\sigma^{(k-1)})h(S_i)$ we have that

$$\log V_i-\log V_{i-1}=\log h(S_i)-\log h(S_{i-1}).$$

Thus the function (3) to be maximized in the k'th iteration of the KMV method can be written as

$$L_V(\mu,\sigma|v_0,\dots v_n)=$$
$$\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2\Delta t}}\exp\left[-\frac{\left(\log h(s_i)-\log h(s_{i-1})-\left(\mu-\sigma^2/2\right)\Delta t\right)^2}{2\sigma^2\Delta t}\right]$$
$$\times\prod_{i=1}^{n}\frac{1}{f(\sigma^{(k-1)})h(s_i)}.$$

Clearly, the maximizer of this function is the maximizer of the first factor. As this does not depend on $\sigma^{(k-1)}$, every iteration will yield the same estimates of μ and σ. Hence, regardless of the initial guess, the KMV algorithm converges in one iteration and to the same value of σ.

Moreover, as $G(s,\sigma) = C^{-1}(s,\sigma)$, we have

$$\frac{1}{C'(v(\sigma),\sigma)} = \frac{1}{\frac{\partial C(v,\sigma)}{\partial v}\Big|_{v=C^{-1}(s,\sigma)}} = \frac{\partial G(s,\sigma)}{\partial s} = f(\sigma)h'(s_i)$$

so that the observed data likelihood (4) can be written as

$$L_S(\mu,\sigma|S_0,\dots S_n) =$$
$$\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2\Delta t}} \exp\left[-\frac{\left(\log h(s_i) - \log h(s_{i-1}) - \left(\mu - \sigma^2/2\right)\Delta t\right)^2}{2\sigma^2\Delta t}\right]$$
$$\times \prod_{i=1}^{n} \frac{h'(s_i)}{h(s_i)}$$

The maximizer of this function is the same as what the KMV method returns after its first iteration, and this proves that in this case, the KMV method yields the MLE estimates. □

This factorization criterion implies that the unobserved market values are proportional to a known function of the observed equity value. The factorization criterion for G is equivalent to $C(v,\sigma) = h^{-1}(v/f(\sigma))$ for some functions f and h, and this is clearly not satisfied for the European call option in the Merton model.

6 | NUMERICAL EVIDENCE

We now present some numerical examples that serve to illustrate our main points. Our examples confirm that the two methods yield different estimates of the volatility, and indicate that the difference seems to be more pronounced when the equity option is far out-of-the-money, that is, when asset value is much lower than the face value of debt. Importantly, even if we start the KMV algorithm at the maximum-likelihood estimate, the algorithm moves away from the starting point to a point of lower likelihood. This illustrates not only that the KMV method does not produce the MLE estimates, but also shows that the KMV method cannot be an EM-algorithm as the likelihood cannot decrease during the KMV algorithm. We return to this more technical issue in the next section.

Our numerical experiment is performed by simulating a time series for the underlying asset value, converting it into equity prices using the Merton model, and treating these simulated equity prices as the observations from which we have to estimate the model parameters. We fix the drift (μ) and the volatility (σ) as well as parameters entering into the option pricing formula, that is, strike price (D), riskless rate (r), and maturity date (T). The parameters in our study are set to $(r, D, \mu, \sigma) = (0.03, 0.8, 0.1, 0.25)$ with an initial time to maturity of $T = 3$ years. The initial value of the underlying assets, V_0, will vary so that we can investigate how moneyness of the equity option (i.e., how high the asset value is compared to the face value of debt) matters for the difference in

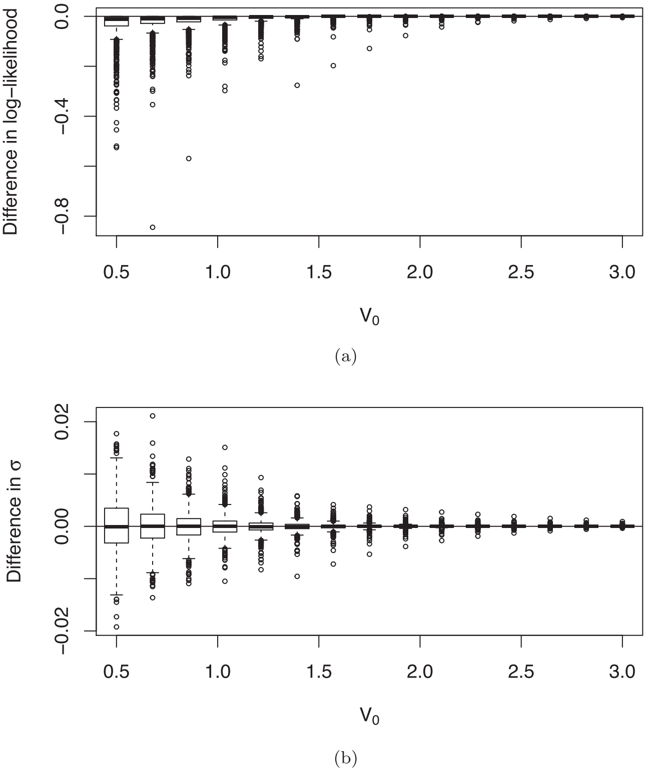


FIGURE 1 The figure shows box plots of differences between log-likelihood values, (a), and estimates of σ, (b), obtained using the KMV method and maximum likelihood in the simulation study with $(r, D, \mu, \sigma) = (0.03, 0.8, 0.1, 0.25)$, a starting time to maturity of $T = 3$ years, and 2 years with 250 observation per year. One thousand simulations are performed for each unique V_0 value

estimates obtained by the two methods. Given a starting value V_0, we simulate the evolution in asset value for 2 years with 250 data points per year, ending up at a time to maturity of 1 year. For each starting value of V_0, we simulate 1000 processes, thereby obtaining a simulated distribution of both estimators for each starting value and the given parameters. Simulated paths in which the asset price, V_i's, drops below 0.01 are excluded as very small asset values can give rise to extremely small equity values when the equity option is deep out-of-the-money, and this may challenge the computational accuracy.

We start both the iterations of the KMV method and the solution of the likelihood equations of the MLE method at $\sigma = 0.1$—that is, some distance away from the true value of 0.25.

Figure 1(a) shows a box plot of differences between the log-likelihood at the MLE estimates and the log-likelihood at the KMV estimates. We observe that unless V_0 is large, there is a clear difference between the log-likelihood at the MLE estimates and KMV estimates, indicating that the KMV method does not give the MLE estimates. To rule out that this is just due to a bad choice of stopping criteria, we have run the KMV method again but starting at the MLE estimates. We always observe that the KMV method in this later run moves away from the MLE estimates and converges to the original estimates of the KMV method, that is, to a point with a lower log-likelihood. Hence the values reported here are valid also for the case where we start the KMV algorithm at the MLE estimates.

To see what this means for the estimates from the two methods, Figure 1(b) shows a box plot for the difference between the estimates of σ derived from the two methods. Again, differences are numerically larger, when V_0 is small. The differences appear to be symmetrically distributed around zero. In particular, this suggests that the estimator derived using the KMV method is as unbiased as the MLE estimates.

TABLE 1 Average differences between quantities in the simulation study with $(r, D, \mu, \sigma) = (0.03, 0.8, 0.1, 0.25)$, a starting time to maturity of $T = 3$ years, and 2 years with 250 observation per year. One thousand simulations are performed for each unique V_0 value. Starting values of the underlying asset, V_0, are shown in the header. Standard errors of the Monte Carlo estimates are shown in the parentheses. Empty cells represent an average difference, which is below the shown precision, that is, smaller than 0.0005.

V_0	0.500	0.857	1.214	1.571	1.929	2.286	2.643	3.000
$\lvert\mu_{\text{MLE}} - \mu_{\text{KMV}}\rvert$	0.004	0.001						
	(0.000)	(0.000)						
$\lvert\sigma_{\text{MLE}} - \sigma_{\text{KMV}}\rvert$	0.004	0.002	0.001					
	(0.000)	(0.000)	(0.000)					
Log-likelihood difference	−0.035	−0.019	−0.008	−0.002	−0.001			
	(0.002)	(0.001)	(0.000)	(0.000)	(0.000)			

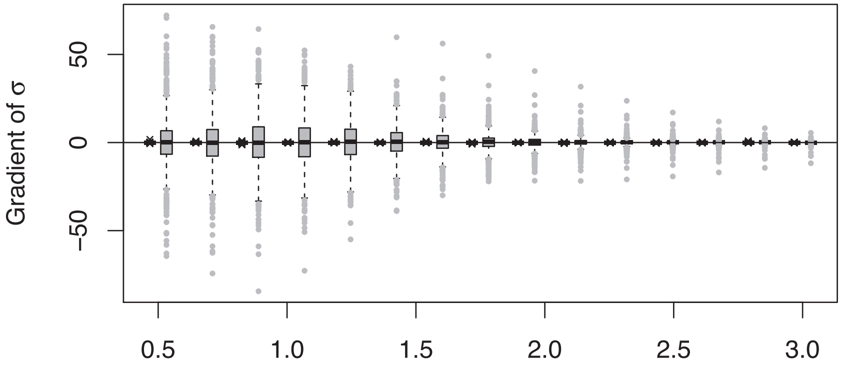
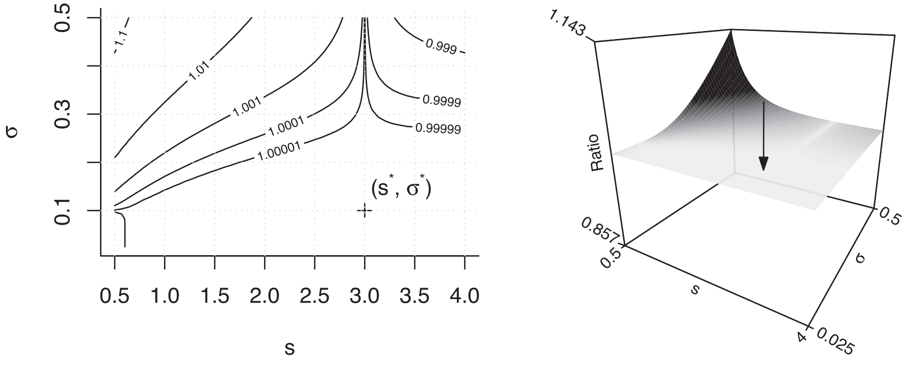


FIGURE 2 The plot shows a box plot of the partial derivative of the log-likelihood with respect to σ at the final estimate in the simulation study with $(r, D, \mu, \sigma) = (0.03, 0.8, 0.1, 0.25)$, a starting time to maturity of $T = 3$ years, and 2 years with 250 observation per year. One thousand simulations are performed for each unique V_0 value. Black crosses and boxes (masked by the crosses) are from the MLE method, and gray circles and boxes are from the KMV method. Both methods use the same simulations and value of V_0 but have been plotted at slightly different positions on the x-axis so that both may be seen

Table 1 further elaborates on this point by showing that the average absolute differences in estimates of drift and volatility are small. The differences vanish for deep in-the-money options. Table 1 also shows that the difference in the log-likelihood is noticeable for fairly high levels of the initial asset value, but there the difference does not produce noticeable differences in the estimates.

Figure 2 shows that the partial derivative of the log-likelihood with respect to σ at the KMV method estimate is clearly nonzero in a large part of the simulations. The graph also shows that the partial derivative is zero at the MLE estimates. This is of course unsurprising, but we include it on the graph to give an idea of the numerical precision of our results. We conclude that the KMV method does not produce the MLE estimates of the unknown parameters.

Finally, to better understand the closeness in estimates when the initial asset value is high, we investigate numerically how big the violation of the factorization condition is for the inverse call option, G, as defined by the Black–Scholes model by inverting Equation (1), that is, whether $G(s, \sigma) \approx f(\sigma)h(s)$ in some region of $C \subseteq (0, \infty)^2$ for some functions f and g. If the condition holds



(a) (b)

FIGURE 3 The figure shows a contour plot in (a) and a 3D plot in (b) of the ratio on the left hand side of Equation (10) evaluated at one (s^*, σ^*) point as a function of (s, σ). This ratio should be one everywhere if the inverse of the option pricing function is multiplicatively separable. The risk free rate is set to $r = 0.03$, the debt is set to $D = 0.8$, and the time to maturity is set to $T = 3$. The levels of the contour lines in (a) differs from one by a factor 10 between each line. The color in (b) is darker the further the ratio is from one. The arrow shows the (s^*, σ^*) point. Both plots show that the ratio is close to one for a large region around (s^*, σ^*)

then for any $(s, \sigma) \in C$

$$\frac{G(s^*, \sigma)G(s, \sigma^*)}{G(s^*, \sigma^*)G(s, \sigma)} \approx 1, \tag{10}$$

for any $(s^*, \sigma^*) \in C$. In Figure 3, we plot the ratio on the left-hand side of Equation (10) for one (s^*, σ^*) point to check if this approximation holds where we set the risk free rate to $r = 0.03$, the debt to $D = 0.8$, and the time to maturity to $T = 3$. Indeed, the figure shows that G seems to approximately factorize for deep in-the-money options with lower values of the volatility. This is consistent with the observation that the two estimation methods yield similar results for values of the underlying which lead to the equity option having a high probability of being deep in-the-money.

7 | THE KMV METHOD IS NOT AN EM-ALGORITHM

Duan et al. (2004) argue that the KMV method is an EM algorithm. This would make the value of the observed data log-likelihood—the log-likelihood based on observation of the S_i's—increase in each iteration and the estimates derived using the EM-algorithm are MLE estimates (if the algorithm is started at a suitable starting value). As we have seen in our numerical experiments, the KMV method moves away from the MLE estimates when started at the MLE estimates and (consequently) decreases the log-likelihood. Hence, despite the apparent similarity between the KMV algorithms and the EM algorithm, Duan et al. (2004)'s argument cannot hold. To explain why, we first review what the EM-algorithm is and why it works.

The EM-algorithm (Dempster et al., 1977) is an algorithm for maximizing the (log-)likelihood in incomplete data situations. It is usually applied in situations, where maximizing the likelihood based on the "complete" data, (X, Y), say, would be simple, but we only observe X. We let $g_\theta(x)$

denote the density of X (i.e., the observed data likelihood) and $f_\theta(x, y)$ the density of (X, Y). Both depend on the unknown parameter, θ.

The EM-algorithm maximizes the observed data log-likelihood, $\log g_\theta(x)$, by iterating two steps, the E-step and the M-step, until convergence. In the $k+1$st iteration, the algorithm first calculates the conditional expectation of the complete data log-likelihood given the observed data using the current estimate $\hat{\theta}^{(k)}$ as the true value of the parameter:

$$Q(\theta|\hat{\theta}^{(k)}) = E_{\hat{\theta}^{(k)}}[\log f_\theta(X, Y)|X = x].$$

This is the E-step. Then, in the M-step, this conditional mean

$$\theta \to Q(\theta|\hat{\theta}^{(k)})$$

is maximized as a function of the unknown parameter θ. The maximizer, $\hat{\theta}^{(k+1)}$, is then the next estimate, which is used in a new E-step followed by an M-step, to get the next estimate, $\hat{\theta}^{(k+2)}$.

Dempster et al. (1977) prove that the value of the observed data log-likelihood is increased (or remains unchanged) in each iteration of the algorithm. We repeat their argument, as this is essential to understanding why the KMV method is not an EM-algorithm:

Theorem 7.1 (Dempster et al. (1977), Theorem 1). *If $(\hat{\theta}^{(k)})_k$ is a sequence of estimators obtained from an EM-algorithm, then for all k*

$$\log g_{\hat{\theta}^{(k+1)}}(x) \geq \log g_{\hat{\theta}^{(k)}}(x).$$

Proof. Write

$$\log g_\theta(X) = \log f_\theta(X, Y) - \log \frac{f_\theta(X, Y)}{g_\theta(X)}, \tag{11}$$

and take conditional expectations (conditioned on $X = x$) to obtain

$$\log g_\theta(x) = Q(\theta|\hat{\theta}^{(k)}) - H(\theta|\hat{\theta}^{(k)}), \tag{12}$$

with

$$H(\theta|\hat{\theta}^{(k)}) = E_{\hat{\theta}^{(k)}}\left[\log \frac{f_\theta(X, Y)}{g_\theta(X)}\middle|X = x\right].$$

As $\hat{\theta}^{(k+1)}$ is obtained by maximizing $Q(\theta|\hat{\theta}^{(k)})$, $\log g_{\hat{\theta}^{(k+1)}}(x) \geq \log g_{\hat{\theta}^{(k)}}(x)$ if

$$H(\theta|\hat{\theta}^{(k)}) \leq H(\hat{\theta}^{(k)}|\hat{\theta}^{(k)}),$$

for all θ. That this is the case follows from Lemma 7.2 below. □

Lemma 7.2 (Dempster et al. (1977), Lemma 1). *For any* θ'

$$H(\theta|\theta') \leq H(\theta'|\theta')$$

for all θ.

Proof. Using Jensen's inequality, we have

$$\begin{aligned} H(\theta'|\theta') - H(\theta|\theta') &= E_{\theta'}\left[\log \frac{f_\theta(X,Y)/g_\theta(X)}{f_{\theta'}(X,Y)/g_{\theta'}(X)}\middle| X = x\right] \\ &\geq \log E_{\theta'}\left[\frac{f_\theta(X,Y)/g_\theta(X)}{f_{\theta'}(X,Y)/g_{\theta'}(X)}\middle| X = x\right] \\ &= \log \int \frac{f_\theta(x,y)/g_\theta(x)}{f_{\theta'}(x,y)/g_{\theta'}(x)} \cdot \frac{f_{\theta'}(x,y)}{g_{\theta'}(x)} dy \\ &= \log \int \frac{f_\theta(x,y)}{g_\theta(x)} dy = 0, \end{aligned}$$

where the last equality follows from the fact that $y \to f_\theta(x,y)/g_\theta(x)$ is a density, namely the conditional density of Y given $X = x$. □

Returning to the KMV algorithm, the observed data (X) are $S = (S_1, \dots, S_n)$ with density given by Equation (4). Duan et al. (2004) let $(S, V) = (S_i, V_i)_{i=1,\dots,n}$ be the complete data, (X, Y), and write the joint density of (S, V) as the density of V given in Equation (3) times a product of indicators ensuring that the S_i's and V_i's have the functional relationship given by Equation (1). As the conditional distribution of V_i given S_i is degenerate, the E-step is easily calculated:

$$\begin{aligned} Q(\sigma|\hat{\sigma}^{(k)}) &= E_{\hat{\sigma}^{(k)}}[\log L_V(\mu, \sigma|V_1, \dots, V_n)|(S_1, \dots, S_n)] \\ &= \log L_V(\mu, \sigma|G(S_1, \hat{\sigma}^{(k)}), \dots, G(S_n, \hat{\sigma}^{(k)})), \end{aligned} \tag{13}$$

where $G(S_i, \hat{\sigma}^{(k)})$ are the "implied asset values." We see, in line with Duan et al. (2004), that the KMV algorithm appears to be an EM algorithm: the E-step calculates "implied asset values" and the M-step finds the next estimate in the KMV algorithm.

To see why the KMV method is not an EM-algorithm, we start with the trivial equality

$$\begin{aligned} \log L_S(\mu, \sigma|S_1, \dots, S_n) = \log L_V(\mu, \sigma|V_1, \dots, V_n) \\ - \log \frac{L_V(\mu, \sigma|V_1, \dots, V_n)}{L_S(\mu, \sigma|S_1, \dots, S_n)}, \end{aligned} \tag{14}$$

similar to Equation (11) in the proof of Theorem 7.1. Using the expressions (3) and (4) and the fact that $V_i = G(S_i, \sigma)$, we may write the last term on the right-hand side as either

$$\sum_{i=1}^{n} \log |C'(G(S_i, \sigma), \sigma)| \quad \text{or} \quad \sum_{i=1}^{n} \log |C'(V_i, \sigma)|. \tag{15}$$

Taking the conditional mean of Equation (14) given $S_1, \dots, S_n$ using $\hat{\sigma}^{(k)}$ as the "true value" of σ does not change the left-hand side of Equation (14) but turns the first term on the right-hand side into $Q(\sigma|\hat{\sigma}^{(k)})$. What happens to the second term on the right-hand side depends on which of the two expressions in Equation (15) we use: The first is unchanged, the second turns into $\sum_{i=1}^{n} \log |C'(G(S_i, \hat{\sigma}^{(k)}), \sigma)|$, but neither of them equals the difference between $\log L_S(\mu, \sigma | S_1, \dots, S_n)$ and $Q(\sigma|\hat{\sigma}^{(k)})$. In other words, the fundamental identity (11) used in the EM-algorithm breaks down. The reason for this is that once we condition using $\hat{\sigma}^{(k)}$ as the "true value" of σ, we break the relationship between V_i and S_i: When σ equals $\hat{\sigma}^{(k)}$, we no longer have $V_i = G(S_i, \sigma)$ and the ratio

$$\frac{L_V(\mu, \sigma | V_1, \dots, V_n)}{L_S(\mu, \sigma | S_1, \dots, S_n)},$$

does not simplify. Instead we could circumvent (14) entirely and define $H(\sigma|\hat{\sigma}^{(k)})$ directly as

$$\begin{aligned} H(\sigma|\hat{\sigma}^{(k)}) &= Q(\sigma|\hat{\sigma}^{(k)}) - \log L_S(\mu, \sigma | S_1, \dots, S_n) \\ &= E_{\hat{\sigma}^{(k)}}\left[\log \frac{L_V(\mu, \sigma | V_1, \dots, V_n)}{L_S(\mu, \sigma | S_1, \dots, S_n)} \middle| S_1, \dots, S_n \right]. \end{aligned}$$

Here, we run into the same paradox as above if we simplify the ratio of the two likelihoods, so we keep in mind that when calculating conditional means using $\hat{\sigma}^{(k)}$ as the parameter V_i does not equal $G(S_i, \sigma)$. Trying to mimic the proof of Lemma 7.2 we trivially have

$$\begin{aligned} &H(\sigma'|\sigma') - H(\sigma|\sigma') \\ &\quad \leq \log E_{\sigma'}\left[\frac{L_V(\mu, \sigma | V_1, \dots, V_n) / L_S(\mu, \sigma | S_1, \dots, S_n)}{L_V(\mu', \sigma' | V_1, \dots, V_n) / L_S(\mu', \sigma' | S_1, \dots, S_n)} \middle| S_1, \dots, S_n \right], \end{aligned}$$

where, of course, the inequality in this case will be an equality, as the conditional expectation is with respect to a degenerate measure. The proof of Lemma 7.2 then proceeds by using that

$$\frac{L_V(\mu, \sigma' | v_1, \dots, v_n)}{L_S(\mu, \sigma' | s_1, \dots, s_n)},$$

is the conditional density of V given $S = s$, we are using to calculate the conditional mean, allowing it to cancel out and leaving us with the "integral" of

$$\frac{L_V(\mu', \sigma | v_1, \dots, v_n)}{L_S(\mu', \sigma | s_1, \dots, s_n)}.$$

Even if we can put some mathematical sense into these ratios as conditional densities (Chang & Pollard, 1997), we would still face the problem that the σ'-conditional distribution is concentrated on another set than the σ-conditional distribution, so there is no mathematical reason why the "integral" would equal 1. Again we see that it is the fact that the relationship between S_i and V_i depends on σ in combination with the conditioning that breaks the argument which would ensure that the observed data log-likelihood is increased in each iteration.

Remark 7.3. As mentioned above, Duan et al. (2004) use (S, V) as "complete data" when arguing that the KMV method is an EM algorithm. For completeness, we note that using (S, V) as "complete data" is mathematically problematic. To understand why this is the case, note that the joint density of S and V could as well have been written as the marginal density of S given by Equation (4) times the same product of indicators as above. But this would give a joint density different from the one used above; these two joint densities are not even proportional to each other as functions of the unknown parameter, thus yielding different complete data MLE estimates. This apparent paradox is measure theoretic: It is caused by using a dominating measure for the densities that is not σ-finite. Without a σ-finite product measure, Tonelli's theorem does not hold, and without Tonelli's theorem, the joint density is not the product of the marginal and the conditional densities. This problem can be fixed by using the unobserved V as the "complete data"; see Chang and Pollard (1997, Example 5, in particular) for mathematical details. Note that using V as the "complete data" amounts to ignoring, as we have already implicity done, the product of indicator functions in the likelihood. This leads to the same expression for $Q(\sigma|\hat{\sigma}^{(k)})$ in Equation (13). In other words, using V as the "complete data" does not change our argument for why the KMV algorithm is not an EM algorithm.

8 | CONCLUSION

The Merton model links the value of a firm's equity to an underlying unobserved asset value. We consider two methods for estimating the parameters driving the unobserved asset value based on observed equity prices. One method is MLE and the other is an iterative approach that to our knowledge was first used by KMV in this context. The two methods produce remarkably similar estimates, but they are not identical, and the difference is most pronounced when asset values are low compared to the face value of the firm's debt, that is, the strike price of the call option, which gives the equity value.

We explain both why the two methods produce different estimates and why they are in some sense close. The first-order conditions for the MLE of the volatility and the fixed point of the KMV method are not identical unless the inverse pricing function is multiplicatively separable. As this is not the case for the inverse of the Black–Scholes function, the two methods produce different results. We provide numerical evidence that the inverse Black–Scholes function is close to being multiplicatively separable for in-the-money options, whereas this is not the case for out-of-the-money options. This explains the often virtual indistinguishability of the estimates produced by the two methods when equity is a deep in-the-money option, whereas there is a larger discrepancy when equity is deep out-of-the-money. We also show that despite a remarkable similarity, the KMV algorithm cannot be viewed as an EM-algorithm.

An argument for using the KMV method is that it is faster. However, the profile likelihood in Equation (7) allows us to find the MLE estimates quickly using fast one-dimensional numerical optimizers. In fact, the difference in computation time between the KMV method and MLE method is not large, when the profile likelihood is used, thus, greatly reducing the benefit of the KMV method.

The KMV method is widely used when the Merton model is implemented empirically and despite the fact that this method produces estimates that are different from the MLE estimates, the differences are so small for the typical levels of firm leverage that the empirical validity is not a problem. However, we also show in our paper that iterative methods like the one proposed by KMV need not converge, and therefore, the strong performance of the KMV algorithm is not something we can expect in general. It would be an interesting area of further research to better understand under what conditions the iterative approach like the KMV method actually converges, and to better understand the properties of the fixed point.

ACKNOWLEDGMENT

Support from the Center for Financial Frictions (FRIC), Grant No. DNRF102, and the Swedish e-Science Research Center are gratefully acknowledged.

ORCID

David Lando https://orcid.org/0000-0003-1118-7096

REFERENCES

Bharath, S., & Shumway, T. (2008). Forecasting default with the Merton distance to default model. *Review of Financial Studies*, *21*, 1339–1369.

Chang, J. T., & Pollard, D. (1997). Conditioning as disintegration. *Statistica Neerlandica*, *51*(3), 287–317.

Crosbie, P., & Bohn, J. (2003). *Modelling default risk*. Moody's KMV.

Dempster, A. P., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, *39*(1), 1–38.

Duan, J. C. (1994). Maximum likelihood estimation using price data of the derivative contract. *Mathematical Finance*, *4*, 155–167.

Duan, J. C., Gauthier, G., & Simonato, J. G. (2004). *On the equivalence of the KMV and maximum likelihood methods for structural credit risk models* [Working paper]. http://neumann.hec.ca/p~240/Recherche/documents/20050617KMVmle.pdf

Merton, R. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, *29*, 449–470.

Vassalou, M., & Xing, Y. (2004). Default risk in equity returns. *Journal of Finance*, *59*(2), 831–868.

How to cite this article: Christoffersen, B., Lando, D., & Nielsen, S. F. (2022). Estimating volatility in the Merton model: The KMV estimate is not maximum likelihood. *Mathematical Finance*, *32,* 1214–1230. https://doi.org/10.1111/mafi.12362