Estimation of Models for Security Prices

Knut K. Aase

Norwegian School of Economics and Business Administration
N-5035 Sandviken
Bergen, Norway

Peter Cутторп

Department of Statistics, CN-22
University of Washington
Seattle, WA 98195, USA

ABSTRACT

In the study of capital markets, security prices play a crucial allocative role. This paper presents stochastic models for the relative security prices, and shows how to estimate these random processes based on historical price data. The models may have continuous components as well as discrete jumps at random time points. Their construction may be a result of a simultaneous equations system, where the underlying determinants are supply and demand. Some of the resulting estimators turn out to be stochastic integrals, which must be computed numerically in practice. Others are of a very simple form, and can be computed directly. Large sample statistical properties, such as consistency and limiting distributions, are derived.
1. Introduction

1.1.

Statistics and probability models have been used in the field of economics for some time now. Few economists deny that uncertainty has an important influence on economic behavior. Almost every phase of consumption and production is affected by uncertainty. Individuals are uncertain about major occurrences in their future lives, such as the epoch of death or their future income. Producers are unsure of their sales and costs. The number of customers arriving at a store, the size of purchases, and the intervals between arrivals are all stochastic. Inventory depletions, equipment breakdowns, depressions and inflations all occur at random time epochs. The results of research and technological processes are probabilistic. One may say that economic agents operate in an environment permeated by stochastic phenomena, and their basic economic decisions are modified accordingly. Consequently, the underlying determinants of supply and demand have stochastic components, and relative prices are random processes.

1.2.

In the study of capital markets, security prices play a crucial allocative role. Firms use them to guide their investment decisions. Similarly, the allocation of an investor's funds across securities is based on these prices. In continuous time, a common stochastic model for security prices has been the geometric Brownian motion. Two classical references to this application are Merton (1971) and Black and Scholes (1973). Newer references include Harrison and Pliska (1981) and Aase (1984a). Whereas the first two works only study processes with continuous sample paths, the other two allow for jumps in the paths as well. In other words, the processes have sample paths that are continu-
ous from the right and have left hand limits (semimartingales). These models are more realistic than the simple diffusion models, that essentially only depend on two parameters. The more general semimartingales constitute a parsimonious class of models, as we shall see, since they can be estimated from historical price data.

The statistical estimation aspect of model building for security prices does not seem to have received attention in the economic literature. It is, however, of considerable practical interest to find good estimates for the local characteristics of the semimartingale price components, since all key decisions are dependent upon these characteristics. The statistical inference is the missing link between the real world and the new advanced probabilistic models. Without it, the semimartingale theory of security decisions is a purely theoretical construct that cannot be applied in practice.

1.3.

This paper deals with statistical estimation of semimartingale models for relative prices of securities. We use likelihood theory to derive estimators of a finite number of parameters. It is also possible to develop nonparametric methods of analysis, but these typically require larger amounts of data, and we postpone the development of such methods to later work. The particular form of the models used in this work has been presented in various economical applications, see e.g. Aase (1983a, 1983b, 1984a) for portfolio optimization problems, Aase (1985) for risky R&D-projects and Aase (1984b) for insurance applications.

The mainstream of statistical inference for models used in economics have occurred in econometrics and time series analysis of economic data. There, stochastic terms are included in the models, since hard empirical facts necessitate such additions. However, they are seldom based in economic theory, and are frequently considered as an inconvenience, which would render superfluous a
properly formulated deterministic model. In these cases, the stochastic terms often are interpreted as resulting from the inadvertent omission of economic variables of importance. Another source of randomness in such models is statistical sampling.

1.4.

The models of the present paper are stochastic, not deterministic. Here the randomness is an integral part of the model, and we do not think of it as a result of omitted variables or sampling. Hence we have to use the recent developments of inference for stochastic processes in order to estimate such models. The models seem to be well suited for fitting security price data, and they take into account fundamental principles of economic theory.

2. The structure of the random model

2.1.

In this section we present the stochastic model for security prices. Given a filtered probability space \((\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})\), we define a system of stochastic differential equations

\[
\frac{dP_i(t)}{P_i(t-)} = \mu_i(t)dt + \sum_{i=1}^d \sigma_i(t)dB_i(t) + \sum_{k=-m}^m \beta_{ik} dN_{ik}(t),
\]

for \(i=1,2,\ldots,d\) and \(t \in [0,T]\). Here \(B(t) = (B_1(t), B_2(t), \ldots, B_d(t))'\) is a \(d\)-dimensional standard Brownian motion with independent components and the \(N_{ik}(t)\) are \(\mathcal{F}_t\)-point processes, counting the number of jumps of size \(\beta_{ik}\) that the relative prices \(P_i(s)/P_i(s-)\) make before time epoch \(t\). Let \(P(t) = (P_1(t), P_2(t), \ldots, P_d(t))'\). We will assume that the jump sizes \(\beta_{ik}, i=1,2,\ldots,d, k=-m,-m+1,\ldots,m-1,m\) are known constants. This assumption is not crucial, but eliminates some technical problems.

Assume the existence of predictable intensities \(\lambda_{ik}(t)\), such that
\( M_{ik}(t) = N_{ik}(t) - \int_0^t \lambda_{ik}(s) ds \) \hspace{1cm} (2.2)

are martingales. (In most applications, predictable means left continuous and \( F_t \)-measurable.) The quantity \( \lambda_{ik}(t) \) can be thought of as the conditional probability of a jump in \( N_{ik} \) at time \( t \), given the history of price data up to that time epoch. In the special case of Poisson point processes, the \( \lambda_{ik} \) are non-random functions. Finally, \( \mu_i(t) \) and \( \sigma_{ij}(t) \) are non-anticipative functionals, i.e. random processes that are adapted to \( F_t \). The matrix \( \sigma \) is, except on the diagonal, the local covariance matrix of the processes \( F_t \), where \( \sigma=(\sigma_{ij}), i,j=1,2,\ldots,d \). For \( i \neq j \) the point processes \( N_{ik} \) and \( N_{jk} \) are assumed orthogonal. Hence the local variance of asset \( i \), \( \sigma^2_i \), is given as

\( \sigma^2_i(t) = \sum_{l=1}^d \sigma^2_{ii}(t) + \sum_{k=-m}^m \beta^2_{ik} \lambda_{ik}(t). \) \hspace{1cm} (2.3)

The process \( P \) is interpreted as the prices of \( d \) securities.

2.2.

The fact that the \( B_i(t), i=1,2,\ldots,d \), are martingales, together with the assumptions made (we use the Ito-interpretation of stochastic integrals with respect to Brownian motion), assures that the efficient market hypothesis holds locally. The terms \( \mu_i(t) \) may be vectors, as may also be the case for \( \beta_{ik}, N_{ik} \) and \( \lambda_{ik} \). Hence (2.1) may be the result of a simultaneous equations model, when this is appropriate. The only additional complications from this are notational, and for simplicity we will confine our attention to the scalar case.

**Example:** A simple special case, similar to one given in Jarrow and Rosenfeld (1984), is the following:

\[
\frac{dP(t)}{P(t-)} = \mu dt + \sigma dB(t) + \sum_{i=1}^q \beta_i dN_i(t).
\]

Here the \( N_i(t) \) are independent Poisson processes of constant rate \( \lambda_i \). A realization of this process is shown in Figure 1. The Jarrow and Rosenfeld model uses a
jump component $\pi dN$ where $\pi$ is a random variable. Our simulation uses the appropriate values for their New York Stock Exchange model, with their random jump size replaced by fixed jump sizes $\beta_i$ corresponding to the quartiles of the Jarrow and Rosenfeld jump distribution. The intuitive interpretation of this model is that relative prices show a linear trend of slope $\mu$, with diffusion fluctuations around the trend line, and occasional, unpredictable jumps of size $\beta_1, \beta_2$ or $\beta_3$.

We shall call the functionals $\mu_i, \sigma_i$ and $\lambda_i$ the local characteristics of the process $P_i(t)$. Given a stretch of data $\{P(t), t \in [0,T]\}$, it is of considerable practical interest to have good estimates of the local characteristics of the semimartingales $P_i(t)$, since they determine the solutions to (2.1) completely. Thus, estimating the characteristics is the same as estimating the model itself.

All key decisions made using these models depend on the values of $\mu, \sigma$ and $\lambda$. For example, any optimal solution to a portfolio dynamic programming problem will depend on them. The problem of statistical estimation of the local characteristics will be the topic of the next section.

3. Parametric estimation based on a likelihood ratio

3.1. Assumptions

We shall assume that the drift components have the following form:

$$\mu_i(t) = \mu_i \varphi_i(t), i = 1, 2, \ldots, d.$$  \hspace{1cm} (3.1)

Here the $\varphi_i$ are known non-anticipative functionals, and the $\mu_i$ are unknown constants, to be estimated from price data. Note the similarity of this assumption to that made in linear simultaneous equations systems. Essentially, we are regressing the drift on the observable processes $\varphi_i$. The variables $\varphi_i(t), 0 \leq t < T$ are observed quantities of economic significance.
We further assume that

$$\lambda_{ik}(t) = \lambda_{ik} \psi_i(t), \ i = 1, 2, \ldots, d, k = -m, -m+1, \ldots, m.$$  

Here the $\psi_i(t)$ are known predictable functionals, and the $\lambda_{ik}$ are unknown constants, again to be estimated from the price data.

We can interpret the coefficients $\lambda_{ik}$ as follows:

$$\lambda_{ik} = P(\text{jump of size } \beta_{ik} \text{ occurs in the relative price process } i \ \mid \ \text{process } i \text{ jumps}).$$

It follows that

$$\sum_{k = -m}^{m} \lambda_{ik} = 1, \ i = 1, 2, \ldots, d \quad (3.2)$$

Reasonable estimators of the $\lambda_{ik}$ will have to satisfy the restriction (3.2). Again, the $\psi_i$ are economically relevant processes, determining the relative rate of jumps at different time epochs.

We now turn to the matrix $\sigma'$. Let us denote the continuous part of the process $P$ by $P^c$. The equation (2.1) can accordingly be written as follows:

$$\frac{dP_i(t)}{P_i(t^-)} = \frac{dP^c_i(t)}{P^c_i(t^-)} + \sum_{k = -m}^{m} \beta_{ik} dN_{ik}(t), \ i = 1, 2, \ldots, d \quad (3.3)$$

We consider continuous sampling throughout the interval $[0, T)$. This means that we can observe $P(t)$ at every instant of time. A consequence of this assumption is that we may consider $\sigma'$ as a known matrix. This follows from Levy's result concerning the quadratic variation of Brownian motion, and its multivariable generalization. This result (cf. Basawa and Prakasa Rao, Lemma 9.4.2) says that

$$\lim_{N \to \infty} \sum_{n=1}^{N} \left[ \frac{P_i((nt 2^{-N}) - P_i((n-1)t 2^{-N}))}{P_i((n-1)t 2^{-N})} \right] \left[ \frac{P_j((nt 2^{-N}) - P_j((n-1)t 2^{-N}))}{P_j((n-1)t 2^{-N})} \right]$$

$$= \sum_{i=1}^{d} \int_{0}^{t} \sigma_i(s) \sigma^T_i(s) ds \ a.s.$$  

for all $t \in [0, T)$. This implies that the matrix $\sigma'$ can be computed with probability one from a single trajectory of the process on any finite interval. In view of this observation there is no loss of generality in assuming $\sigma'$ completely known whenever continuous sampling is involved. Even if we observe the process
discretely at a fine grid of time epochs, the above limit will be a good approximation.

Thus we can confine our attention to the problem of estimating the $\mu_i$'s and the $\lambda_k$'s.

3.2. The likelihood ratio

In order to find the maximum likelihood estimators of the unknown parameters we proceed as follows. Let

$$ dp^c = \left[ \frac{dP^c_1}{P_1}, \frac{dP^c_2}{P_2}, \ldots, \frac{dP^c_d}{P_d} \right] $$

and let $(x,y) = \sum_{i=1}^{d} x_i y_i$ be the inner product of the two vectors $x$ and $y$. For each $t \geq 0$ we define

$$ L_t = 1 + \int_0^t \left\{ (\sigma \sigma')^{-1} \mu(s), dp^c(s) \right\} + \sum_{i=1}^{d} \sum_{k=-m}^{m} (\lambda_k(s)-1) dM_{i,k}(s). \tag{3.4} $$

By use of the Doléans-Dade formula (an extension of Girsanov's exponential formula, see Elliott, 1982), we find that

$$ L_t = \exp\left[ \int_0^t \left( \mu(s), dp^c(s) \right) - \frac{1}{2} \int_0^t \left( \sigma \sigma', \mu(s) \right) ds \right] \tag{3.5} $$

$$ + \sum_{i=1}^{d} \sum_{k=-m}^{m} \ln \lambda_k(s) dN_{i,k}(s) + \int_0^t \sum_{i=1}^{d} \sum_{k=-m}^{m} (1-\lambda_k(s)) ds \right\}. $$

It follows from Lemma 7 in Kabanov et al. (1979) that

$$ L_t = E\left( \frac{dP}{dP_0} \mid F_t \right) $$

and

$$ L_T = \frac{dP}{dP_0}, $$

where $P_0$ is a reference measure. The measure $P_0$ can be taken as a solution to (2.1) with $\mu_i(t) = 0$ and $\lambda_k(t) = 1$ for all $i$ and $k$.

Notice that from (3.4) and (3.5) we obtain the correct expression for the likelihood in the special cases of Ito-processes and of point processes separately.
(see e.g. Taraskin, 1974 and Bremaud, 1981). The fact that the mixed process has the given density follows from the additive form of the semimartingale $M_t$, given in equation (104) of Kabanov et al. (1979), where

$$L_t = 1 + \int_0^t L_s \, dM_s.$$ 

We are now in a position to find the maximum likelihood estimators of the unknown parameters, by maximizing $L_T$ with respect to these parameters, subject to the constraint (3.2).

3.3. The estimators of $\lambda_{ik}$.

We first consider the problem of estimating the parameters $\lambda_{ik}$. We have to maximize

$$F(\lambda, \Lambda) = \log L_T - \sum_{i=1}^d \Lambda_i \left( \sum_{k=-m}^m \lambda_{ik} - 1 \right),$$

where the $\Lambda_i$ are Lagrange multipliers. Setting the partial derivatives equal to zero, we get

$$\hat{\lambda}_{ik} = \frac{N_{ik}(T)}{p \int_0^T \psi_i(t) \, dt + \Lambda_i}$$

and

$$\Lambda_i = \sum_{k=-m}^m N_{ik}(T) - \int_0^T \psi_i(t) \, dt.$$ 

Hence

$$\hat{\lambda}_{ik} = \frac{N_{ik}(T)}{p \sum_{i=-m}^m N_{ii}(T)}, \quad i=1,2, \ldots, d, \quad k=-m, -m+1, \ldots, m$$

are the maximum likelihood estimators of the $\lambda_{ik}$, satisfying the constraint (3.2). Notice that they are extremely easy to compute, and do not depend on the functionals $\psi_i(t), 0 \leq t < T$. Hence the assumption of known $\psi_i$ is inessential. It will be seen later that the estimates of the $\mu_i$ also are free of $\psi_i$, but that the standard
error of the $\hat{\lambda}_k$ does depend on $\psi_i$.

3.4. Statistical properties of $\hat{\lambda}$

The main properties of maximum likelihood estimators of parametrized conditional intensities can be found in Sagalovsky (1982). Since $N_{ik}(t)$ has stochastic intensity $\lambda_{ik}\psi_i(t)$, we see (using Sagalovsky's Corollary 1) that

$$\frac{N_{ik}(T)}{\int_0^T \psi_i(t) dt} \xrightarrow{a.s.} \lambda_{ik} \text{, provided } \int_0^T \psi_i(t) dt \to \infty.$$  

Thus

$$\hat{\lambda}_{ik} = \frac{N_{ik}(T)/\int_0^T \psi_i(t) dt}{\sum_{i=-m}^m N_{ii}(T)/\int_0^T \psi_i(t) dt} \xrightarrow{a.s.} \frac{\lambda_{ik}}{\sum_{i=-m}^m \lambda_{ii}} = \lambda_{ik}.$$  

It is slightly more complicated to find the asymptotic distribution of $\hat{\lambda}_{ik}$. Let $I_{ik}(t) = \lambda_{ik} \int_0^t \psi_i(u) du$ be the integrated intensity, and define its inverse by $I_{ik}^{-1}(t) = \inf \{u : I_{ik}(u) \geq t\}$. Sagalovsky's Theorem 3 can be used to show that

$$\sqrt{I_{ik}(T)} \left[ \frac{N_{ik}(T)}{\int_0^T \psi_i(t) dt} - \lambda_{ik} \right] \xrightarrow{d} N(0, \lambda_{ik}^{-1}) \text{ as } \delta \to \infty.$$  

This is a form of inverse sampling: instead of waiting a specified amount of time, we wait until the integrated intensity has reached a particular value. For practical purposes we will say that $\hat{\lambda}_{ik}$ is approximately normally distributed, with mean $\lambda_{ik}$ and variance $\frac{\lambda_{ik}}{\int_0^T \psi_i(t) dt}$. Thus the variability of $\hat{\lambda}_{ik}$ is intimately connected to $\psi_i(t)$. However, we may estimate the asymptotic variance by $N_{ik}(T)/(\sum_{i=-m}^m N_{ii}(T))^2$, since $E\sum_{i} N_{ii}(T) = E\int_0^T \psi_i(t) dt$.  

3.5. The estimators of $\mu$

Denote the (known) elements of the matrix $[\sigma\sigma]^\dagger$ by $(a_{ij})$. Notice that $a_{ij}=a_{ji}$, since $(a_{ij})$ is the inverse of a covariance matrix. A straightforward computation shows that

$$
\frac{\partial}{\partial \mu_i} \ln L_T = \int_0^T \varphi_i(t) \sum_{j=1}^d a_{ji} dp_j(t) - \frac{1}{2} \sum_{j=1}^d \mu_j \int_0^T (\varphi_i(t) + \varphi_j(t)) a_{ij} dt.
$$

Also

$$
\frac{\partial^2}{\partial \mu_i \partial \mu_j} \ln L_T = -\frac{1}{2} \int_0^T (\varphi_i(t) + \varphi_j(t)) a_{ij} dt. \tag{3.8}
$$

The negative of the $(dxd)$-matrix in (3.8) will be denoted $B_T$. We can write the likelihood equations in compact form as

$$
b_T - B_T \hat{\mu} = 0 \tag{3.9}
$$

where $b_T$ is the $(dx1)$-vector with components

$$
\int_0^T \varphi_i(t) \sum_{j=1}^d a_{ji} dp_j(t), \ i=1,2,\ldots,d.
$$

Since these equations are linear in the $\hat{\mu}_j$, we can solve (3.9) by Cramer's rule if certain invertibility conditions are met. When this is the case, we find the maximum likelihood estimator of $\mu$ as

$$
\hat{\mu} = B_T^{-1} b_T. \tag{3.10}
$$

$b_T$ is a stochastic integral which is evaluated numerically over an increasingly fine grid using McKeans' (1970) result that

$$
\int_0^T \varphi_i(t) \sum_{k=1}^d a_{ek} dp_k(t) = \lim_{N \to \infty} \sum_{j=1}^{2^N} \varphi_i((j-1)t 2^{-N}) \sum_{k=1}^d a_{ek} (p_k(jt 2^{-N}) - p_k((j-1)t 2^{-N})) \text{ a.s.}
$$

3.6. Statistical properties of $\hat{\mu}$

Our basic assumption in this section is that

$$
\frac{1}{T} B_T \xrightarrow{p} C \text{ as } T \to \infty \tag{3.11}
$$

where the elements $c_{ij}$ of the non-singular matrix $C$ satisfy $|c_{ij}| < \infty$ (see also Aase, 1982). Under this assumption, the following result can be shown.
For any $\varepsilon > 0$ there is a $T_\varepsilon > 0$ such that for any $T > T_\varepsilon$, with probability at least $1 - \varepsilon$, the likelihood equation (3.9) has a unique solution $\hat{\mu}_T$ which is a consistent estimator of the vector $\mu$, i.e.

$$\hat{\mu}_T \overset{a.s.}{\rightarrow} \mu$$

Furthermore, the distribution of $\sqrt{T}(\hat{\mu} - \mu)$ is asymptotically normal with mean zero and covariance matrix $C^{-1}$. For details, see e.g. Basawa and Prakasa Rao (1980, section 9.5) or Aase (1982). Thus, we treat $\hat{\mu}_T$ as approximately normally distributed with mean $\mu$ and covariance matrix $(B_T)^{-1}$.

In the case where

$$\mu_i(t) = \mu_i \varphi_i(P(t)), \quad i = 1, 2, \ldots, d$$

and

$$\sigma(t) = \sigma(P(t)), \quad \text{i.e. when we have time homogeneity and no aftereffects (Markovian structure).}$$

we can show an additional result. It hinges on the assumption that the continuous part of the process has an ergodic distribution $G_\mu(P)$ for each $\mu$. In that case the asymptotic covariance matrix takes the following simple form:

$$\left[ \frac{1}{2} \int_0^T (\varphi_i(P) + \varphi_j(P)) \alpha_{ij} \, dG_\mu(P) \right]^{-1}. \quad (3.11)$$

On the other hand, if $\mu_i(t)$ and $\sigma(t)$ only depend on time (i.e. are non-random), one can show that $\hat{\mu}$ is an efficient estimator, and that the vector $\sqrt{T}(\hat{\mu} - \mu)$ has a normal distribution with mean zero and covariance matrix $B_T^{-1}$. Notice that in this case $B_T$ is a non-random matrix. For details, we refer again to Basawa and Prakasa Rao (1980, section 9.5). For alternative recursive estimators see Aase (1982).

Acknowledgements:

Part of this work was done while the first author was visiting the University of Washington. The second author had partial support from the National Science
Foundation under grants MCS-8205991 and MCS-8302573.

References:


Figure caption

Figure 1. A simulated path for logarithms of (relative) prices at the New York stock exchange.