Likelihood Based Surveillance of Continuous-Time Processes

Helgi Tomasson

Address: Helgi Tomasson  
Faculty of Economics and Business Administration  
University of Iceland  
Oddi, at Sturlugata, 101 Reykjavik  
Iceland

Email: helgito@hi.is
Likelihood Based Surveillance of Continuous-Time Processes.

Helgi Tómasson

July 2, 2007


# 1 Introduction

In every day life individuals and firms are exposed to a constant flow of data. An example is the news broadcast on modern television, a person is reading the news and at the bottom of the screen there is a banner with a constant flow of quotes of a financial market. Information on time of transactions, prices and volume are observed and stored electronically by systems such as Reuter and Bloomberg. It is of interest to form objective tools to do inference based on such streams. To assess the informative value of such a data stream a suitable statistical model is required together with a matching statistical inference strategy. Firms, like banks, financial supervisory authorities, central-banks and other financial institutions are facing a similar data-stream. Firms, buying, selling or distributing financial products need to keep track of the content of data, and update their prices accordingly. These institutions need an objective tool as a basis for their decision. The aim of this chapter is to suggest an approach of monitoring changes in a continuous-time price process.

In order to develop an objective tool some formal definitions are necessary. Due to the academic success of mathematical finance in recent years, many of the individuals making decisions are aware of pricing rules such as Black-Scholes, etc. The Black-Scholes rule is an example of a popular way of formalizing dynamics with help of stochastic differential equations (SDE). The mathematical pricing rules, such as Black-Scholes, are based on parameters in the SDE and the institutions decide on a change in their prices based on their inference about an eventual parameter shift in the SDE. The SDE describes a continuous process, but in practice the observed data are collected
discretely, often at uneven time intervals. An example of such data flow is the quotes from financial markets.

In statistical subcultures there are many approaches for the phenomenon of a parameter shift. Broemeling & Tsurumi (1987) give the following classification for the varying parameter problem (change-of-regime, structural-break etc.). First, the case of a known break-point and an abrupt change in parameter. Second, an abrupt change in parameter at an unknown time point. Third, that there is a gradual (non-stochastic) change in the parameter over a certain period. Fourth, the parameters might follow a stochastic process. And fifth, data might be from a mixture of populations. In this chapter, only breaks of the second type are considered. The concept of parameter constancy, often in terms like, structural-change, regime-shift is treated in many text books on econometrics. Article collections on statistical analysis of economic structural change are given in Hackl (1989) and Hackl & Westlund (1991). In this chapter the focus is on on-line detection of a deterministic (exogenous) abrupt parameter change. The methodological approach is based on ideas of statistical surveillance. The concepts of surveillance are reviewed in Frisén (2003).

In this chapter, an approach of implementing likelihood-based surveillance tools to continuous-time diffusion processes is given. In Chapter 2 some background to diffusion processes is given. The regime-shift is supposed to be of the type, a jump in a parameter. The dynamics of the process are of
the form:

\[ dX(t) = \mu(X(t), t) \, dt + \sigma(X(t), t) \, dW(t), \]

\[ \mu(X(t), t) = \mu_1(X(t), t)I_{[t<\tau]} + \mu_2(X(t), t)I_{[\tau<t]}, \]

\[ \sigma(X(t), t) = \sigma_1(X(t), t)I_{[t<\tau]} + \sigma_2(X(t), t)I_{[\tau<t]}. \]

Popular continuous-time financial models for interest rates are single factor mean-reverting processes. Even though they look simple, data-analysis in these simple continuous-time models is somewhat problematic because important features, such as the transition density is not known in closed form. Recently, A"ıt-Sahalia (1999, 2002) has in a couple of articles suggested a Taylor-approximation of the transition density and thereby opened the computational possibility of working numerically with the likelihood function. This allows the possibility of likelihood inference, i.e. maximum-likelihood estimation, Bayesian estimation and likelihood-based surveillance. In this chapter on-line monitoring a simple continuous-time model is addressed. Data are supposed to be generated by a simple mean-reverting process, the CKLS, (Chan, Karolyi, Longstaff & Sanders, 1992),

\[ dX(t) = \kappa(\alpha - X(t))dt + \sigma X(t)^\rho dW(t) \]  

which contains the popular CIR (Cox, Ingersoll & Ross, 1985), \( \rho = 1/2 \), as a special case. Schmid & Tzotchev (2004) describe an approach for surveillance of a CIR model. In this chapter the CKLS model is treated as an illustrative example. The reasons for that choice are perhaps the same as for its popularity in interest rate analysis; it is a simple model that can capture several stylized facts of interest rate movements. It is always positive, it is asymmetric around its mean and it can have heavy tails. The parameters
have somewhat an easy interpretation. The $\alpha$ parameter is the long-term mean. The $\kappa$ parameter controls the speed of convergence to that mean. The $\sigma$ and $\rho$ parameters are linked to properties such as variance and tail behaviour. The CKLS family is, as most diffusions, analytically difficult. Therefore traditional statistical analysis, like maximum-likelihood estimation has to rely on numerical approaches. In Section 2 a brief review of the problems of calculating the likelihood function is given. The approximations of the likelihood function are done by means of Taylor expansion. A somewhat more detailed description is given in appendix A. For understanding of the ability to detect a break it is necessary to have an idea about the precision of the estimates of the parameters and the nature of the model. In Section 3 some numerical properties of the CKLS model are discussed and some examples of maximum-likelihood estimates shown. Having an algorithm for calculating the likelihood functions offers the possibility of applying well known likelihood-based surveillance tools such as (exponential)-CUSUM and the Shiryaev-Roberts (SR) statistics. In this chapter the CUSUM and SR statistics for dynamic continuous-time processes are implemented in the notation and spirit of Shiryaev (2002). Even though the dynamics are specified in continuous-time, data are assumed to consist of discrete observations and thus also decisions are made in discrete time. The simulation of a continuous-time process and calculation of the likelihood therefore require some discretization tools. The implementation of the discretization tools is described in Section 4. An illustration of implementation and computation is given in Section 5. The calculations are performed using the statistical environment, R, and R-packages written by the author. A version of the R-
surveillance-package for the CKLS model may be obtained from the author.

2 Likelihood approximations for diffusion processes

The diffusion-process in the form of a stochastic differential equation (SDE) is written in the form:

\[ dX(t) = \mu(X(t), t) \, dt + \sigma(X(t), t) \, dW(t). \]

The concept of a diffusion process is briefly reviewed in Chapter 2 of this volume. In this case only one-dimensional time-homogeneous SDE’s are considered. The drift and diffusion term are assumed to be on the form \( \mu(x, \theta) \) and \( \sigma(x, \theta) \) where \( \theta \) is a vector of parameters.

The log-likelihood process, \( \Lambda(T) \), for a continuous-time diffusion process observed in the time interval \( [0, T] \) is of the form:

\[
\log(\Lambda(T)) = c + \int_0^T \frac{\mu(X(t), t)}{\sigma(X(t), t)^2} \, dX(t) - \frac{1}{2} \int_0^T \frac{\mu(X(t), t)^2}{\sigma(X(t), t)^2} \, dt.
\] (2)

Maximizing this expression, (2), with respect to the parameter vector \( \theta \) will yield the maximum-likelihood estimator, \( \theta_{ML} \), which will (in general) be a function of the entire path of \( X(t) \) in the time interval \( [0, T] \). For some special process analytical solutions of the maximization of (2) are available, (Kutoyants, 2004). In most cases it is not possible to write (2) in closed form, let alone find a simple analytical form of the maximum-likelihood estimator. In practical cases the process, \( X(t) \), is only observed at discrete time-points,
A discretised version of (2) is:

\[
\sum_{i=2}^{T} \mu(X(t_{i-1}), t_{i-1})^2 (X(t_i) - X(t_{i-1})) - \frac{1}{2} \sum_{i=2}^{T} \sigma(X(t_{i-1}), t_{i-1})^2 \Delta_i,
\]

\[
\Delta_i = t_i - t_{i-1}.
\]

(3)

Equation (3) represents a particular approximation of the true likelihood. For calculating the true likelihood of a discretely observed diffusion it is necessary to have a way to calculate its transition density, \( f(X(t + \Delta)|X(t)) \). The transition density of a diffusion-process \( X(t) \) is the density function of random-variable \( X(t + \Delta) \) conditioned on its value \( x_0 = X(t) \) at time \( t \). For most diffusions no analytical form of the transition density is known. The idea of Aït-Sahalia (1999, 2002) is to utilize the fact that the transition-density, \( f(x|x_0, \Delta) \) of a diffusion satisfies the Kolmogorov forward equation,

\[
\frac{\partial f(x|x_0, \Delta)}{\partial \Delta} + \frac{\partial (\mu(x, \theta)f(x|x_0, \Delta))}{\partial x} - \frac{1}{2} \frac{\partial^2 (v(x, \theta)f(x|x_0, \Delta))}{\partial x^2} = 0,
\]

(4)

where \( v(x) = \sigma^2(x) \). The technique is based on doing a Taylor expansion of \( f(x|x_0, \Delta) \) in \( \Delta \) around 0. Instead of working with \( f \) directly it is possible to work with \( l(x|x_0, \Delta) = \log(f(x|x_0, \Delta)) \). That way it is ensured that the approximation of the density will stay positive. Substituting \( \exp(l(x|x_0, \Delta)) \) into equation (4) we see that \( l(x|x_0, \Delta) \) will satisfy a new partial differential equation:

\[
\frac{\partial l(x|x_0, \Delta)}{\partial \Delta} + \mu'(x) + \mu(x) \frac{\partial l(x|x_0, \Delta)}{\partial x}
\]

\[
- \frac{1}{2} \left( \frac{\partial l(x|x_0, \Delta)}{\partial x} \right)^2 - \frac{1}{2} \frac{\partial^2 l(x|x_0, \Delta)}{\partial x^2} = 0.
\]

(5)
Then \( l(x|x_0, \Delta) \) is Taylor expanded around \( \Delta = 0 \),

\[
l(x|x_0, \Delta) = -\frac{1}{2} \log(2\pi \Delta) - \frac{1}{2} \log(v(x)) + \frac{c_{-1}(x|x_0)}{\Delta} + \frac{c_0(x|x_0) + c_1(x|x_0)\Delta + c_2(x|x_0)\Delta^2}{2!} + \cdots + \frac{c_k(x|x_0)\Delta^k}{k!} + R_k(\Delta, x|x_0).
\]

Substituting (6) into (5) and matching powers of \( \Delta \) gives a system of differential equations. A brief review of the system is given in appendix A.

Each set of functions \( \mu(x) \) and \( \sigma(x) \) have their Taylor expansion. If a set of observations of the diffusion process, \( x(t_1), x(t_2), \ldots, x(t_n) \) is obtained then the log-likelihood can be approximated with \( l^* \), and the optimization problem:

\[
\max_{\theta} l^*(\theta | X(t_1), \ldots, X(t_n)),
\]

is solved with numerical methods. The quality of the approximation depends on the number of terms in the expansion, \( \Delta \), \( (x - x_0) \), and some properties of \( \mu \) and \( \sigma \).

**Remark.** There exist many equivalent parameterizations of the diffusion processes. For example a Box-Cox transformation of the CKLS process has unit diffusion coefficient. If \( X(t) \) is CKLS and \( Y(t) \) is defined as:

\[
Y(t) = \frac{1}{\sigma} \frac{X(t)^{1-\rho}}{1-\rho}.
\]

Then the diffusion \( Y(t) \) will have unit diffusion with drift:

\[
\mu_Y(y) = \frac{\kappa}{\sigma} \left( ((\rho - 1)\sigma y)^{\frac{1}{1-\rho}} \right)^{-\rho} \left( \alpha - ((\rho - 1)\sigma y)^{\frac{1}{1-\rho}} \right) - \frac{1}{2} \rho \sigma \left( ((\rho - 1)\sigma y)^{\frac{1}{1-\rho}} \right)^{\rho-1} - \rho\sigma y^{\frac{\rho}{1-\rho}}.
\]

A process with unit diffusion coefficients may be easier to deal with, e.g. with non-linear least squares software, but the interpretation of the parameters in (7) is less transparent than in the standard CKLS representation. In the
CKLS case the Box-Cox transformation, aimed to make the process “more normal” is an example of a transformation that converts the diffusion function into a constant. Such a transformation is available for all one-dimensional diffusion processes. See, appendix A.

3 Illustration of numerical properties

Results by Jensen & Poulsen (2002), Lindström (2004) and Aït-Sahalia (1999, 2002) show that estimation methods based on the Taylor approximations described in Section 2 seem to perform favorable to alternative methods for many simple models used in the financial literature. A popular class of models is the class of the mean-reverting models. The idea is that there is a force that pushes the series towards an equilibrium value. Some further details are described in Chapter 2. A simple form is:

\[ dX(t) = \kappa(\alpha - X(t)) \, dt + \sigma(X(t)) \, dW(t), \]

which has a solution:

\[ X(t) = x_0 + \int_{t_0}^{t} \kappa(\alpha - X(s)) \, ds + \int_{0}^{t} \sigma(X(s)) \, dW(s). \]

(8)

It is easy to show that the conditional expectation of \( X(t) \) given by equation (8) is:

\[ E(X(t)|X(t_0) = x_0) = \alpha + \exp(-\kappa t)(x_0 - \alpha). \]

(9)

The interpretation is straightforward, \( \alpha \) is the long-term equilibrium and \( \kappa \) describes the speed of convergence to this equilibrium. The half-time of a deviance from equilibrium in equation (9) is \( \log(2)/\kappa \). The function \( \sigma(x) \) controls the stochastic impact. For some simple forms of \( \sigma(x) \) it is possible
to derive a closed form of the conditional variance $V(X(t)|X(t_0) = x_0)$. In general the CKLS model of equation (1) not much can explicitly written down except for the mean. Analytically, $\rho = 1/2$ and $\rho = 1$ are a little bit easier. For $\rho = 1/2$, the CIR-model, the theoretical likelihood function is known in closed form. For $\rho = 1$ it is possible to calculate conditional mean and variance, i.e., $E(X(t)|X(s))$ and $V(X(t)|X(s))$ for $t > s$.

The process $X(t)$ is sampled at time-points $t_1, \ldots, t_n$, generates data on the form:

$$x(t_1), \ldots, x(t_n), \quad t_1, \Delta_2 = t_2 - t_1, \ldots, \Delta_n = t_n - t_{n-1}.$$ 

In traditional statistical textbooks consistent estimates are obtained by letting $n \to \infty$. In the continuous-time environment things are more complicated and it is needed that both $\Delta_i \to 0$ and $t_n - t_1 \to \infty$. Letting $n \to \infty$, but say, keeping $t_n - t_1 = 1$ would give consistent estimates on $\sigma(x)$ but not on the drift, and similarly, if all the $\Delta_i$’s are large, the estimate of the diffusion function $\sigma(x)$ is likely to be poor.

To illustrate the usefulness of the Taylor-approximation of the likelihood and the sensitivity to time-span and sampling interval a brief simulation study was performed. The Taylor expansions mentioned in Section 2 were programmed in FORTRAN, and the maximization of the likelihood was performed numerically with the R-optimization routine `optim` which allows constraints on the parameters. The CKLS process of equation (1) was simulated using the Milstein-scheme, (Kloeden & Platen, 1992). The parameter values were constrained to realistic values, $\kappa, \alpha, \sigma$ were bounded away from zero, $\rho$ was bounded away from $1/2$. The setup of the simulation was as follows:

- The time spans used are $T= 1, 10, 100$. 
• $\Delta=1, 0.1$ and $0.01$ are used.

• 10 points of process per observation were simulated by the Milstein scheme.

• $\kappa = 0.24, \alpha = 0.07, \sigma = 0.08838, \rho = 0.75$. (Values from CKLS and others)

Results based on 25 replications are shown in Tables 1 through 4. It is clear from the tables that the quality of parameter estimates in the drift function, $\kappa$ and $\alpha$ increases with the length of the time interval $T$. It is also clear from the tables that the quality of the parameters in the diffusion function, $\sigma$ and $\rho$ increases with increased sampling frequency, smaller $\Delta$. A sample of size 100, $\Delta = 1$ and $T = 100$ is more informative about $\kappa$ than an sample of size 100 with $\Delta = 0.01$ and $T = 1$.

<table>
<thead>
<tr>
<th></th>
<th>T=1</th>
<th>T=10</th>
<th>T=100</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\kappa}$</td>
<td>3.2742</td>
<td>0.5352</td>
<td>0.2709</td>
</tr>
<tr>
<td>$\hat{\alpha}$</td>
<td>0.0793</td>
<td>0.0962</td>
<td>0.0695</td>
</tr>
<tr>
<td>$\hat{\sigma}$</td>
<td>0.1115</td>
<td>0.0979</td>
<td>0.0899</td>
</tr>
<tr>
<td>$\hat{\rho}$</td>
<td>0.7695</td>
<td>0.7732</td>
<td>0.7570</td>
</tr>
</tbody>
</table>

Table 1: Average estimates, for $\Delta=0.01$

Just as in ARMA and many other models, identifiability is an issue. In the general CKLS it will be hard to distinguish between $\rho$ and $\sigma$ if both are large. An intuitive explanation is as follows. If we have a diffusion of the above form then the longterm average is $0.07$. Then the diffusion term is
\[ \Delta=1 \quad \Delta=0.1 \]

<table>
<thead>
<tr>
<th></th>
<th>( \hat{\kappa} )</th>
<th>( \hat{\alpha} )</th>
<th>( \hat{\sigma} )</th>
<th>( \hat{\rho} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Delta=1 )</td>
<td>0.8232 0.2916 0.9842 0.8642 0.7342 0.7299</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \Delta=0.1 )</td>
<td>0.0644 0.0744 0.0542 0.0309 0.2794 0.1084</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Average estimates, for \( T=100 \)

\[ \begin{array}{cccc}
\text{T=1} & \text{T=10} & \text{T=100} \\
\text{s.d. } \hat{\kappa} & 1.8703 & 0.3995 & 0.0693 \\
\text{s.d. } \hat{\alpha} & 0.0234 & 0.1269 & 0.0048 \\
\text{s.d. } \hat{\sigma} & 0.0542 & 0.0309 & 0.0042 \\
\text{s.d. } \hat{\rho} & 0.2794 & 0.1084 & 0.0176 \\
\end{array} \]

Table 3: Standard deviation of simulations, for \( \Delta=0.01 \)

frequently close to:

\[ \sigma 0.07^\rho (\frac{x}{0.07})^\rho. \]

If \( \rho \) is increased and \( x \) close to its mean, the effect on the diffusion function rarely affected if \( \sigma \) is increased accordingly. In more complex diffusion models it seems reasonable that this issue will become more complicated.

Under certain conditions there exists an invariant stationary distribution of the diffusion. When it exists, its density is on the form:

\[ f(x) \propto \frac{1}{\sigma^2(x)} \exp\left( \int_c^x \frac{2\mu(s)}{\sigma^2(s)} ds \right). \]

For the case \( \rho = 1/2 \), it is possible to write down the transition density. The
stationary distribution will be a gamma distribution:

\[ f(x) \propto x^{2\alpha/\sigma^2 - 1} \exp\left(\frac{2\kappa}{\sigma^2}x\right), \]

with mean \( \alpha \) and variance \( \frac{\sigma^2}{2\kappa} \).

The parameter \( \kappa \) controls the speed of convergence to the mean, a low value means slow convergence and therefore a large unconditional variance. If \( 2\kappa\alpha/\sigma^2 > 1 \) the CIR process will never hit zero, the invariant distribution will be a gamma distribution and all moments of exist. The case for \( \rho > 1/2 \) is slightly different and it will not be possible to write down a closed form transition density. The case \( \rho = 1 \) is analytically convenient. For that case the invariant distribution will be:

\[ f(x) \propto x^{2\kappa/\sigma^2 - 2} \exp\left(-\frac{2\kappa\alpha}{\sigma^2 x}\right). \] (10)

The distribution in equation (10) is the inverse-gamma distribution. If \( 2\kappa\alpha/\sigma^2 > 0 \) it has mean \( \alpha \) and if \( 2\kappa/\sigma^2 > 1 \) it will have a variance. For large values of \( x \), part \( B \) in equation (10) will be close to 1, and the behavior of the density will be dominated by part \( A \) which is a Pareto-type tail. The existence of moments is therefore analogous to the Pareto distribution. Due
to this fact, for some values of the parameters only a few moments might exist. Therefore extreme observations are likely to occur and due to the dynamic properties of the process, observations close in time to the extreme observation are likely to extreme as well.

When $1/2 < \rho < 1$ all moments of the invariant distribution exists, but if $\kappa$ is small, the behavior of the process can be quite similar to a heavy-tail case. For example, if $\rho = 3/5$ then the invariant distribution is:

$$f(x) \propto \frac{1}{\sigma^2 x^{6/5}} \exp\left(-\frac{5\kappa(4\alpha + x)}{2\sigma^2 x^{1/5}}\right).$$

(11)

Part $B$, the exponential part, in equation (11) ensures the existence of moments. However, for a large range of $x$-values the impact of the exponential part is very little. Say, if $\alpha = 0.05$, then most of the time the $X(t)$ process stays in low values. If $\kappa$ is low the effect of the exponential part of (11) is very little and part $A$ dominates the behavior of the invariant distribution. In this particular case the term, $A$, is similar to a (heavy-tailed) Pareto distribution. Therefore if $\kappa$ is very small extreme values are likely to be somewhat persistent. As illustration a simulation experiment with $\kappa = 0.01, \alpha = 0.05, \sigma = 0.09, \rho = 0.6$ was performed. A plausible path of 4000 days (time-units) is illustrated in figures 1 and 2. The for long periods the process stays below 0.02 and even below 0.01, i.e. corresponding to interest rates of 1%-2%. Occasionally it climbs up to 30-40%. In the period between day 500 and day 1000 it rockets off to several hundred percent, i.e. a factor hundred times its usual value. When a long horizon (4000 days) is viewed on the last graph on figure 2, it seems that the extreme period was short. From the above, it is clear that the CKLS can generate a large variety of paths.
Figure 1: A short horizon view of a semi-heavy tail process. The top figure shows the first 100 days, the middle one day 101 through 200 and the bottom one shows a 20 day period from day 151 through 170.

Figure 2: A long horizon view of a semi-heavy tail process, The top figure shows the first 500 days, the middle one day 500 through 1000, and the bottom one shows days 1 through 4000.
4 Implementation of surveillance tools

The tools of numerical approximation of the likelihood function of a diffusion process offers the possibility of applying standard surveillance tools, such as the CUSUM and Shiryaev-Roberts. Here the notation of Shiryaev (2002) for a continuous-time setup is used. Shiryaev (2002) and Srivastava & Wu (1993) give analytical results for the case where there is a change at time $\tau$ in the drift of a Wiener process. This turns out to be an analytically tractable case,

$$X(t) = r(t - \tau)I_{[\tau, \infty]}(t) + \sigma W(t),$$
$$dX(t) = rI_{[\tau, \infty]}(t) + \sigma dW(t).$$

where $W(t)$ is a standard Wiener process (Brownian motion). Following the notation of Shiryaev (2002), two probability-measures, $P_\tau$ and $P_\infty$ are to be compared. The probability measures represent a shift in the model at time $\tau$ and a model with no shift ($\tau = \infty$), respectively. The system is observed up to time $t$. The principal tool is the likelihood-ratio process:

$$L(t) = \frac{dP_\tau}{dP_\infty}(t, X(t)),$$

where $\frac{dP_\tau}{dP_\infty}$, the Radon-Nikodym derivative. The CUSUM statistic in continuous time, is:

$$\log(\gamma(t)) = \log(\max_{\tau \leq t} \frac{L(t)}{L(\tau)}).$$ (12)

The Shiryaev-Roberts (SR) statistic in continuous time is:

$$\psi(t) = \int_0^t \frac{L(t)}{L(\tau)} \, d\tau.$$ (13)
In this case it is easily derived by substituting into the log-likelihood-process formula, equation (2), that the log-likelihood-ratio process, for shift at \( \tau = 0 \), is:

\[
log(L(t)) = \frac{r}{\sigma^2} X(t) - \frac{r^2}{2\sigma^2} t. 
\]

In general \( log(L(t)) \) will be a function of an integral involving the whole path of \( X(t) \). The processes \( \gamma(t) \) and \( \psi(t) \) are in general dynamic processes with some dependency structure. In the case of the Wiener process, only the value \( X(t) \) and some deterministic function of \( t \) enter formula of the likelihood-ratio process, which makes analytical derivations possible. By the use of Ito’s formula Srivastava & Wu (1993) give the dynamics of \( \psi(t) \) in the Wiener process case:

\[
d\psi(t) = dt + \frac{r}{\sigma^2} \psi(t) dX(t).
\]

Srivastava & Wu (1993) show that the distribution of the CUSUM process, \( log(\gamma(t)) \), is in the case of a shift of drift in a Wiener process, the same as of \( |Z(t)| \) where:

\[
dZ(t) = \frac{r}{2} \text{sgn}(t - \tau)\text{sgn}(Z(t))dt + dW(t),
\]

\text{sgn} is the sign operator.

In the Wiener process with drift case, both the CUSUM-process and the SR-process can therefore be described by a stochastic differential equation. To the author’s knowledge dynamic forms of \( \gamma(t) \) and \( \psi(t) \) are not known explicitly for other cases. In a recent article by Baron & Tartakovsky (2006) it is stated that “for continuous-time models, beyond the problem of detecting a change in the drift of a Brownian motion, little is known about the properties
of CUSUM and SR procedures”. Baron & Tartakovsky (2006) also review some asymptotic optimality questions on the change-point detection problem.

In this article the focus is on empirical methods and computations. In practical cases the process, $X(t)$, is observed at discrete time-points, $t_1, \ldots, t_n$. Here, no particular sampling strategy is assumed. Srivastava & Wu (1994) consider the possibility of dynamic sampling in the case of a shift in drift of a Wiener process. In financial data it is also conceivable that the sampling time-points are generated by separate trading process. Here the trading process is assumed to exogenous. In the notation of Shiryaev (2002) the discretized version of equations (12) and (13), for sampling points, $t_1, t_2, \ldots, t_n$, are:

$$
\log(\gamma(t_k)) = \log(f_1(x(t_1), \ldots, x(t_k))) - \min_{0 \leq \tau \leq t_k} (\log(f_0(x(t_1), \ldots, x(\tau))))
$$

(14)

and

$$
\psi(t_k) = \sum_{j=1}^{k} \frac{f_1(x(t_1), \ldots, x(t_j))}{f_0(x(t_1), \ldots, x(t_j))} (t_k - t_{k-1}),
$$

(15)

where $f_0$ and $f_1$ are the densities before and after time $\tau$, respectively. Frisén (2003) gives some optimality properties of these statistics. The alarm-rules are: Alarm if $\gamma(t_k) > \text{constant}$ or $\psi(t_k) > \text{constant}$. Alarm rules will be of the form, alarm at time $\tau_A$ such that:

$$
\tau_A = \inf\{t > 0; \gamma(t) > d\} \quad \text{or} \quad \tau_A = \inf\{t > 0; \psi(t) > d^*\}.
$$

The constants, $d$, and $d^*$, depend on which models are compared, the duration between measurements ($\Delta_i$’s), and the false alarm strategy. Illustration
Figure 3: Some examples of a shift from model $M_0$ to $M_1$ at $\tau = 1000$.

of some numerical properties of the maximum-likelihood estimator given in Section 3 shows that the amount of information about the parameters in the drift function is primarily a function of the time span, whereas information about parameters in the diffusion function increases when density of observation increases. It is therefore, intuitively clear that the nature of increased sampling frequency will affect characteristics of the alarm statistics $\gamma(t)$ and $\psi(t)$.

The control-limits, $d$ and $d^*$ have to be decided according to some alarming principles like $ARL_0$. A way is to use quantiles of $\gamma(t)$ and $\psi(t)$ or choose a particular length of $ARL_0$. These depend on the distributions of the alarm statistics, $\gamma(t)$ and $\psi(t)$, that are not available in closed form so, they have to be decided by simulations.
5 Calculation of alarm statistics

A nice feature of the CKLS is the easy of interpretation of parameters. $\kappa$ is the speed of convergence to the mean, $\alpha$, $\sigma$ is (partly) a scale parameter and $\rho$ (partly) rules the tail behaviour of the distribution. For $\rho < 1$ all moments of the invariant distribution exist but only some in the case $\rho \geq 1$. A lot of quantitative finance focuses on estimation of variance, but also tail-probabilities are of interest. As seen in a previous section, CKLS can generate some interesting heavy-tail patterns.

A small simulation study for illustrating the behaviour of the SR and CUSUM surveillance tools was performed and some of the results are shown in this section. The alarm rules depend on the parameter values of the competing models as well as the sizes of the sampling intervals. To illustrate the nature of this dependence some numerical experiments were performed.
A base model, $M_0$, was set up:

$M_0 : \theta = (\kappa = 0.1, \alpha = 0.05, \sigma = 0.1, \rho = 0.75)$

versus a shift in $\rho$

$M_1 : \theta = (\kappa = 0.1, \alpha = 0.05, \sigma = 0.1, \rho = 1.5)$

versus a shift in $\kappa$

$M_2 : \theta = (\kappa = 0.05, \alpha = 0.05, \sigma = 0.1, \rho = 0.75)$

These models were chosen to contrast the behaviour due to a shift in a parameter in the drift function to a shift of parameter in the diffusion function. The shift of $\rho = 0.75$ to $\rho = 1.5$ represents a shift from a light-tail invariant distribution to a heavy tailed one. The shift of $\kappa$ from 0.1 to 0.05 describes the situation of that the half-life of a deviance from the longterm mean ($\alpha = 0.05$) goes from $\log(2)/0.1 \simeq 7$ to $\log(2)/0.05 \simeq 14$, so if time is measured in days, this means that the half-length of a cycle goes from one week to two weeks. Examples of shifts from $M_0$ to $M_1$ and $M_2$, respectively, are shown in Figures 3 and 4. The shift from $M_0$ is in some sense a more drastic one. The equilibrium distribution shifts from being a light-tailed on to being a heavy-tailed one. Riskmanagers, would like to avoid the heavy-tailed case. The shift from $M_0$ to $M_2$ is a more moderate case. Financial analyst sometimes claim that bubble size is increasing. Analysing cases such as $M_2$ is a way of formalizing such a statement.

For a practical analysis of surveillance of a continuous-time process important factors are the sampling frequency and the time-span. To illustrate the behaviour of $\gamma(t)$ and $\psi(t)$ after a break has occurred, a simulated process was generated with the models $M_1$ and $M_2$, respectively, and the hypothetical parameter shift supposed to take place at time $\tau = 0$. The Milstein-scheme
was used to generate 100 points per day (time-unit). Two types of sampling were compared, two observations per day, $\Delta = 0.5$, and ten observations per day $\Delta = 0.1$.

Of course the true $\gamma(t)$ and $\psi(t)$ cannot be observed as they are functions of the entire sample path of $X(t)$ and $X(t)$ is only observed at discrete time-points. Therefore the estimates $\hat{\gamma}(t)$ and $\hat{\psi}(t)$ are calculated based on equations (14) and (15). In this section these estimates will be denoted as, $\gamma(t)$ and $\psi(t)$. The quality of the estimates will depend on $\Delta$ and the models compared. Numerically the estimate of $\psi(t)$, based on equation (15) can be zero, especially when no break takes place. Therefore, the graphs illustrating the no-change case in this section $\log(\psi(t) + 1)$ is plotted instead of $\log(\psi(t))$.

In Figures 5 and 6 the processes $\log(\gamma(t))$ and $\log(\psi(t) + 1)$ are plotted over a period of 10 years. In Figure 5 the data are generated by $M_0$ and CUSUM and SR calculated by reference to $M_1$. In Figure 6 the same data are generated by $M_0$ and CUSUM and SR are calculated by reference to $M_2$. From the figures it is apparent that the CUSUM and SR tend to agree on the alarm. As seen in Figures 5 and 6, both $\log(\gamma(t))$ and $\log(\psi(t)) + 1$ seem to be mean-variance stationary (as expected) in the case of no parameter-shift.

In Figure 6 there are clear signs of auto-correlation in both CUSUM and SR. If the state of alarm is defined by a high value of CUSUM or SR, then the system will be in the state of alarm for days when monitoring $\kappa$, whereas when monitoring $\rho$ the alarm periods are very short, e.g. one day. If the system has been out of alarm for a day, the waiting time for the next alarm is empirically similar to an exponential distribution for the case of $\rho$.

In contrast to the no-change cases described in Figures 5 and 6, there
will be an upward trend in $\log(\gamma(t))$ and $\log(\psi(t))$, when $M_0$ is no longer the true model. The CUSUM and SR process seem therefore to be behaving as they should. Except for the Wiener process with drift case, there are not any closed form results for the distributional behaviour of CUSUM and SR, therefore, for exact evaluation dynamic properties, $ARL_0$, quantiles, etc., of $\gamma(t)$ and $\psi(t)$, extensive simulation is needed.

The results are as to be expected. In Figures 7 and 8 $\log(\psi(t))$ and $\log(\gamma(t))$ are plotted for 16 replications of the case when $M_1$ is the model and $M_0$ the reference model. In all 16 cases there is a clear upward trend and comparing to Figure 5 a decisive alarm is available in a few days. A small $\Delta$ will result in a quicker detection of change. For both CUSUM and SR the trend is approximately, 3 units per day for $\Delta = 1/2$ and 15 units per day for $\Delta = 1/10$. That is, an increase of sampling frequency by factor 5 increases accumulation of information by approximately factor 5. In contrast it will take a long time to detect a shift in $\kappa$. An increased sampling frequency adds very little information about $\kappa$. An intuitive explanation is that a value of $\kappa$ of the size 0.1 can easily generate a bubble (cycle) of say 50-100 days. During such a period very little evidence is gained about an eventual change of cycle length, by increasing the sampling frequency. Therefore, to be able to detect a change in cycle (bubble) length, many cycles have to be observed, and it takes time to collect sufficiently many bubbles. In Figures 9 and 10 $\log(\gamma(t))$ and $\log(\psi(t))$ are plotted for a single replication of the case where $M_2$ is the true model and $M_0$ is the reference model. The impact of the size of $\Delta$ is virtually invisible. In this particular case the Figures, 9 and 10, show that it takes roughly 100-200 days before evidence against $M_0$ in favour of $M_2$ really
starts to pile up. For each realization of true model $M_2$ versus $M_0$, CUSUM and SR behave somewhat like a step function indicating that there are long periods that are non-informative about $\kappa$. This behaviour was confirmed in other replications. This is natural as change in bubble length can not be detected if no bubbles take place.

![CUSUM and SR plots](image.png)

Figure 5: CUSUM=$\log(\gamma(t))$ and log(SR)=$\log(\psi(t)+1)$ for true model $M_0$ versus $M_1$. 
Figure 6: CUSUM=$\log(\gamma(t))$ and log(SR)=$\log(\psi(t)+1)$ for true model $M_0$ versus $M_2$.

Figure 7: Effects of sampling density on $\log(\psi(t))$. 16 replications of true model $M_1$ against $M_0$ for $\Delta = 1/2$ and $\Delta = 1/10$. 
Figure 8: Effects of sampling density on $\log(\gamma(t))$. 16 replications of true model $M_1$ against $M_0$ for $\Delta = 1/2$ and $\Delta = 1/10$.

Figure 9: Effects of sampling frequency on $\log(\text{SR})=\log(\psi(t))$ for a shift from model $M_0$ to $M_2$ at $t = 0$. 
Figure 10: Effects of sampling frequency on CUSUM for a shift from model $M_0$ to $M_2$ at $t = 0$. 
6 Discussion

The approximations of the likelihood function of a diffusion process offer the possibility of a surveillance of parametric stability in CKLS type processes and also a large class of other univariate diffusions. Many of the popular mathematical finance models can be monitored with methods of this type. The parameters of a CKLS model are interpretable and it is important that the analysts understand their nature. It is inherent in the CKLS model that it will be difficult to estimate the speed of convergence to an equilibrium, $\kappa$, when $\kappa$ is a small number. It will also in some cases be difficult to separate the parameters, e.g. $\rho$ and $\sigma$. This puts an extra responsibility on the analyst. The analyst has to make sensible restrictions on the parameters, and make preferences which features are of interest. The Taylor approximations of the likelihood together with modern computer programs on maximization and numerical integration offer a computational procedure for likelihood-based surveillance of continuous-time processes. Having a numerical approximation to the likelihood functions offers the possibility to apply standard surveillance tools such as CUSUM and SR. These tools have some optimality properties, see, Frisén (2003) and Baron & Tartakovsky (2006), but analytical results on how to calculate their dynamics have been limited. It seems evident that the properties of CUSUM and SR for CKLS models, such as worst-case scenarios, average-run-length, distribution of waiting-time for alarm, etc., will have to be evaluated by simulation methods. The properties will depend on the parameter values as well as the sampling frequency. It seems that monitoring parameters in the drift function is difficult; nothing can replace a long period of observation, whereas monitoring parameters in the diffusion function can
be improved by denser sampling of the process. That is good news, because from a financial point of view, risk-management, option-pricing, hedging, the diffusion function is of vital importance. There are few analytical results available for likelihood analysis of continuous-time finance. Therefore, approximations and extensive simulations will be necessary. The popularity of mathematical finance has had the impact that there are now many individuals in the finance industry that have the knowledge and understanding of diffusion models. For them interpretation of parameters in a diffusion models is natural. A statistical tool for linking data to continuous-time modelling allows direct inference about parameters of interest in an particular model.
Appendix

Taylor expansions

The coefficient of $\Delta^{-1}$, $c_{-1}(x|x_0)$ solves

$$2c_{-1}(x|x_0) + v(x)(c'_{-1}(x|x_0))^2 = 0$$

$$2\mu(x)c'_{-1}(x|x_0) - 2v(x)c'_0(x|x_0)c'_{-1}(x|x_0)$$

$$-v'(x)c'_{-1}(x|x_0) - v(x)c''_{-1}(x|x_0) - 1 = 0$$

This gives:

$$c_{-1}(x|x_0) = -\frac{1}{2} \left( \int_{x_0}^{x} \frac{du}{\sqrt{v(u)}} \right)^2$$

$$c_0(x|x_0) = \int_{x_0}^{x} (2\mu(s)c'_{-1}(s) - c'_{-1}(s|x_0)v'(x) - v(s)c''_{-1}(s|x_0) - 1)/(2v(s)c_{-1}(s|x_0))ds$$

The remaining equations, $j = 1, 2, \ldots$, to be solved are of the type:

$$c_j(x|x_0) + c'_j(x|x_0)k_j(x) = g_j(x)$$

Where:

$$k_1(x) = -v(x)c'_{-1}(x)$$

$$k_2(x) = -\frac{1}{2}v(x)c''_{-1}(x)$$

$$k_3(x) = -\frac{1}{3}v(x)c'''_{-1}(x)$$

$$k_4(x) = -\frac{1}{4}v(x)c''''_{-1}(x)$$
and

\[ g_1(x) = -\mu(x)c_0(x) - \mu'(x) - \frac{1}{8} \frac{v'(x)^2}{v(x)} + \frac{1}{2} v(x)c_0'(x) + \frac{1}{2} \]
\[ v'(x)c_0'(x) + \frac{1}{2} v(x)c_0''(x) + \frac{1}{4} v''(x) + \frac{1}{2} \frac{\mu(x)qqv'(x)}{v(x)} \]
\[ g_2(x) = -\mu(x)c_1'(x) + v(x)c_0'(x)c_1'(x) + \frac{1}{2} v'(x)c_1'(x) + \frac{1}{2} v(x)c_1''(x) \]
\[ g_3(x) = -\mu(x)c_2'(x) + v(x)c_0'(x)c_2'(x) + v(x)c_1'(x) + \frac{1}{2} v'(x)c_2'(x) + \frac{1}{2} v(x)c_2''(x) \]
\[ g_4(x) = -\mu(x)c_3'(x) + v(x)c_0'(x)c_3'(x) + 3v(x)c_1'(x)c_3'(x) + \frac{1}{2} v'(x)c_3'(x) + \frac{1}{2} v(x)c_3''(x) \]
\[ g_5(x) = -\mu(x)c_4'(x) + v(x)c_0'(x)c_4'(x) + 3v(x)c_2'(x) + \frac{1}{2} v'(x)c_4'(x) + \frac{1}{2} v(x)c_4''(x) \]
\[ g_6(x) = -\mu(x)c_5'(x) + v(x)c_0'(x)c_5'(x) + 5v(x)c_1'(x)c_5'(x) + 10v(x)c_2'(x)c_3'(x) + \frac{1}{2} v'(x)c_5'(x) + \frac{1}{2} v(x)c_5''(x) \]
\[ \vdots \]

From elementary calculus it is well known that:

\[ c_j(x|x_0) = e^{-A(x)} \int_{x_0}^{x} \frac{g_j(s)}{k_j(s)} e^{A(s)} ds \quad (1) \]
\[ A(x) = \int_{x_0}^{x} \frac{1}{k(s)} ds \quad (2) \]

Solving these differential equations sequentially by integrating (1) can in some cases be problematic. A working solution is to Taylor expand \( c_j(x|x_0) \) in \( x \) around \( x_0 \) and substitute the Taylor expansion instead of the functions \( g_j \). A trick that is sometimes useful is to rewrite the diffusion such that
For a one-dimensional diffusion this is always possible by using the transformation \( y = g(x) \). defined by.

\[
Y(t) = \int_{X_0}^{X(t)} \frac{du}{\sigma(u)}
\]

Then by Ito’s lemma the dynamics of the process \( Y(t) \) is described with:

\[
dY(t) = \mu_Y(Y(t)) \, dt + dW(t)
\]

\[
\mu_Y(x) = \frac{\mu_X(x)}{\sigma(x)} - \sigma'(x) \quad \text{with} \quad x = g^{-1}(y)
\]

For example if we have a CIR process:

\[
dX(t) = \kappa (\alpha - X(t)) \, dt + \sigma \sqrt{X(t)} \, dW(t)
\]

\[
g(x) = \int_{x}^{\infty} \frac{du}{\sigma \sqrt{u}} = \frac{2}{\sigma} \sqrt{x}, \quad g^{-1}(y) = \frac{\sigma^2 y^2}{4}
\]

\[
dY(t) = \mu_Y(Y(t)) \, dt + dW(t)
\]

\[
\mu_Y(y) = \frac{4 \kappa \alpha - \sigma^2}{2 \sigma^2 y} - \frac{\kappa y}{2}
\]

This transformation makes calculations of the Taylor expansion and computer programming simpler, but it can also cause numerical difficulties in some cases. In practical cases both transformed and untransformed calculations should be done for validation. For more details consult Aït-Sahalia (1999) and Aït-Sahalia (2002).
References


Editor Sveinn Agnarsson

A complete list of IoES Working Papers and Reprints can be accessed on our World Wide Web site at http://www.ioes.hi.is


W07:05 Anna Gunnthorsdottir, Roumen Vragov, Kevin McCabe and Stefan Seifert: The meritocracy as a mechanism to overcome social dilemmas.

W07:04 Gylfi Zoega: Endogenous employment cycles in Euroland.

W07:03 Thorvaldur Gylfason: The international economics of natural resources and growth.

W07:02 Helga Kristjansdottir: Talking trade or talking aid? Does investment substitute for aid in the developing countries?


W06:13 Brynhildur Davidsdottir: Sustainable energy development.

W06:12 Helga Kristjansdottir: Substitution between inward and outward foreign direct investment.

W06:11 Helga Kristjansdottir: Evaluation of Icelandic trade flows, the gravity model approach.

W06:10 Brynhildur Davidsdottir: Capital constraints and the effectiveness of environmental policy.

W06:09 Gylfi Zoega: Market forces and the continent’s growth problem.

W06:08 Fridrik M Baldursson and Nils-Henrik M von der Fehr: Vertical integration and long-term contracts in risky markets.

W06:07 Ragnar Arnason: Conflicting uses of marine resources: Can ITQ’s promote an efficient solution?

W06:06 Thorvaldur Gylfason and Gylfi Zoega: A golden rule of depreciation.

W06:05 Ron Smith and Gylfi Zoega: Global factor, capital adjustment and the natural rate.

W06:04 Thorolfur Matthiasson: To whom should the rent accrue?

W06:03 Tryggvi Thor Herbertsson and Gylfi Zoega: Iceland’s Currency Dilemma.

W06:02 Thorolfur Matthiasson: Possible stakeholder conflicts in quota regulated fisheries, contribution to the political economics of fisheries.

W06:01: Eyjolfur Sigurdsson, Kristin Siggeirsdottir, Halldor Jonsson jr, Vilmundur Gudnason, Thorolfur Matthiasson, Brynjolfur Y Jonsson: Early discharge and home intervention reduces unit costs after total hip replacement: Results of a cost analysis in a randomized study.

W05:14 Gylfi Zoega and J Michael Orszag: Are Risky Workers More Valuable to Firms?

W05:13 Friðrik Mári Baldursson: Fairness and pressure group competition.
W05:12 Marias H. Gestsson and Tryggvi Thor Herbertsson: Fiscal Policy as a Stabilizing Tool.

W05:11 Tryggvi Thor Herbertsson and Gylfi Zoega: On the Adverse Effects of Development Aid.

W05:10 Thráinn Eggertsson and Tryggvi Thor Herbertsson: Evolution of Financial Institutions: Iceland’s Path from Repression to Eruption.


W05:08 Ron Smith and Gylfi Zoega: Unemployment, investment and global expected returns: A panel FAVAR approach.

W05:07 Gylfi Zoega and Thorlakur Karlsson: Does Wage Compression Explain Rigid Money Wages?

W05:06 Thorvaldur Gylfason: India and China

W05:05 Edmund S. Phelps: Can Capitalism Survive?

W05:04 Thorvaldur Gylfason: Institutions, Human Capital, and Diversification of Rentier Economies

W05:03 Jón Danielsson and Ásgeir Jónsson: Countercyclical Capital and Currency Dependence

W05:02 Alison L. Booth and Gylfi Zoega: Worker Heterogeneity, Intra-firm Externalities and Wage Compression

W05:01 Tryggvi Thor Herbertsson and Martin Paldam: Does development aid help poor countries catch up?

W04:12 Tryggvi Thor Herbertsson: Personal Pensions and Markets

W04:11 Fridrik M. Baldursson and Sigurdur Johannesson: Countervailing Power in the Icelandic Cement Industry

W04:10 Fridrik M. Baldursson: Property by ultimatum: The case of the Reykjavik Savings Bank

W04:09 Ingólfur Arnarson: Analyzing Behavior of Agents of Economic Processes in Time

W04:08 Otto Biering Ottosson and Thorolfur Matthiasson: Subsidizing the Icelandic Fisheries


W04:06 Ingolfur Arnarson: Modelling Fishery Management Schemes with an Olympic System Application

W04:05 Ingolfur Arnarson and Pall Jenson: Adding the Sales Markets Dimension to Bio-Economic Models. The Case of Fishery Management

W04:04 Edmund S. Phelps: Changing Prospects, Speculative Swings: Structuralist Links through Real Asset Prices and Exchange Rates

W04:03 Ingolfur Arnarson: Analysing Behavior of Agents of Economic Processes in Time

W04:02 Ron Smith and Gylfi Zoega: Global Shocks and Unemployment Adjustment

W04:01 Fridrik M. Baldursson and Nils-Henrik M von der Fehr: Prices vs. quantities: public finance and the choice of regulatory instruments
W03:07 Sveinn Agnarsson and Ragnar Arnason: The Role of the Fishing Industry in the Icelandic Economy. A historical Examination

W03:06 Thorolfur Matthiasson: Paying paper by paper, the wage system of Icelandic University teachers explained

W03:05 Gur Ofur and Ilana Grau: Bringing the Government hospitals into line: The next step of reform in the healthcare sector

W03:04 Ingolfur Arnarson and Pall Jennisson: The Impact of the Cost of the Time Resource on the Efficiency of Economic Processes

W03:03 Torben M. Andersen and Tryggvi Thor Herbertsson: Measuring Globalization

W03:02 Tryggvi Thor Herbertsson and J. Michael Orszag: The Early Retirement Burden: Assessing the Costs of the Continued Prevalence of Early Retirement in OECD Countries

W03:01 Eirik S. Amundsen, Fridrik M. Baldursson and Jørgen Birk Mortensen: Price Volatility and Banking in Green Certificate Markets

W02:10 Tryggvi Thor Herbertsson and Gylfi Zoega: A Microstate with Scale Economies: The Case of Iceland

W02:09 Alison, L. Booth and Gylfi Zoega: Is Wage Compression a Necessary Condition for Firm-Financed General Training

W02:08 Asgeir Jonsson: Exchange rate interventions in centralized labor markets

W02:07 Alison, L. Booth, Marco Francesconi and Gylfi Zoega: Oligopsony, Institutions and the Efficiency of General Training

W02:06 Alison L. Booth and Gylfi Zoega: If you’re so smart, why aren’t you rich? Wage inequality with heterogeneous workers