On exact solutions of some changepoint detection problems

Mikhail Zhitlukhin

Steklov Mathematical Institute, Moscow

(jointly with Albert Shiryaev and William Ziemba)

Perethof, 18 July 2017

Introduction

This talk gives an overview of some recent results in statistical methods of detecting changes in random sequences and processes.

Example 1

A random sequence $X_n = \xi_1 + \ldots + \xi_n$, where ξ_k are independent and their distribution changes from F to G at time θ:

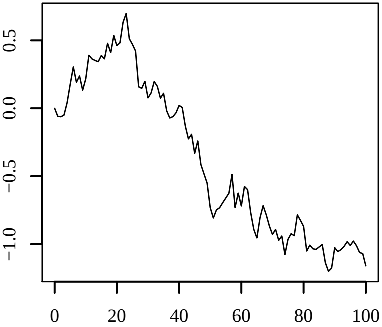
$$\xi_k \sim F \text{ for } k < \theta, \quad \xi_k \sim G \text{ for } k \geqslant \theta.$$

Example 2

A Brownian motion with drift appearing after time θ:

$$X_t = \mu(t-\theta)^+ + B_t.$$

An example of a changepoint:



Overview of the talk

1. Brief history
2. A general changepoint model and sufficient statistics
3. Corollaries: solutions of some particular problems
4. Some financial applications

Remarks

– In this talk only on-line detection procedures are considered: they make a decision about the change in a non-anticipative way.

– We'll focus on exact optimal detection rules, mostly in Bayesian settings

History of changepoint detection theory

- One of the first who considered changepoint detection problems was W. A. Shewhart: the method of control charts (1925)

- The basic statistics for simple alternatives and iid observations were studied by Roberts, Page, Shiryaev, Lorden and others in the 1960-70s

- Further developments: multiple post-change alternatives, non-iid observations, combining with optimal control problems, etc.
  A good review can be found in Tartakovsky & Nikiforov (2015)

A general changepoint model

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0})$ be a filtered measurable space and P, Q are two locally equivalent probability measures ($\mathrm{P}_t \sim \mathrm{Q}_t$ for all $t \geqslant 0$):

- P describes the behavior before the change,
- Q after the change.

We want to paste these two measures together at some time $\theta \geqslant 0$.

Example

$\Omega = C([0,\infty))$ and ω_t is a Brownian motion under P and a Brownian motion with drift μ under Q. Then $d\mathrm{Q}_t = \exp(\mu\omega_t - \frac{\mu^2}{2}t)d\mathrm{P}_t$.
The measure P^θ corresponding to the change at time θ is defined by

$$\frac{d\mathrm{P}_t^\theta}{d\mathrm{P}_t} = \begin{cases} \exp(\mu(\omega_t - \omega_\theta) - \frac{\mu^2}{2}(t-\theta)), & t \geqslant \theta, \\ 1, & t < \theta. \end{cases}$$

Let $\mathrm{R} = (\mathrm{P} + \mathrm{Q})/2$ so that $\mathrm{P}, \mathrm{Q} \ll \mathrm{R}$, and define the density processes

$$Z_t^0 = \frac{d\mathrm{Q}_t}{d\mathrm{R}_t}, \qquad Z_t^\infty = \frac{d\mathrm{P}_t}{d\mathrm{R}_t}, \qquad t \geqslant 0.$$

Let Z_t^θ be the new density process

$$Z_t^\theta = Z_t^\infty \mathbf{I}(t < \theta) + Z_t^0 \frac{Z_{\theta-}^\infty}{Z_{\theta-}^0} \mathbf{I}(t \geqslant \theta)$$

of the corresponding measure P^θ such that for all $A \in \mathcal{F}_t$

$$\mathrm{P}^\theta(A) = \mathrm{E}^{\mathrm{R}}(Z_t^\theta \mathbf{I}(A)).$$

We interpret $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geqslant 0}, \mathrm{P}^\theta)$ as the model with a changepoint at θ.

Criteria of optimality of changepoint detection rules

A changepoint detection rule is identified with a stopping time τ of the filtration $\mathcal{F}_t$. We want τ to be as close as possible to θ in some sense, for example minimizing $\mathrm{E}|\tau - \theta|$.

A Bayesian problem

Let $G(t)$ be a distribution function on $\mathbb{R}_+$. Consider the following optimality criterion:

$$\int_0^\infty \mathrm{E}^\theta|\tau - \theta| dG(\theta) \to \min .$$

Then θ can be thought of as a random variable with prior distribution G.

(Below we'll also discuss a wider class of penalty functions instead of $|\tau - \theta|$)

Namely, define on the space $(\Omega \times \mathbb{R}_+, \mathcal{F} \otimes \mathcal{B}(\mathbb{R}_+))$ the measure

$$\mathrm{P}^G(A \times B) = \int_B \mathrm{P}^t(A) dG(t).$$

Then the process $X_t((\omega_1, \omega_2)) = \omega_1(t)$ can be interpreted as the observable process, and the random variable $\theta((\omega_1, \omega_2)) = \omega_2$ as the change-point.

The optimality criterion above can be written as follows:

$$\int_0^\infty \mathrm{E}^\theta|\tau - \theta| dG(\theta) = \mathrm{E}^G|\tau - \theta|.$$

Introduce the process (the so-called Shiryaev–Roberts statistic):

$$\psi_t = \int_0^t \frac{Z_t^s}{Z_s^\infty} dG(s).$$

Proposition 1. We show that

$$\min_\tau \int_0^\infty \mathrm{E}^\theta|\tau - \theta| dG(\theta) = \min_\tau \mathrm{E}^{\mathrm{P}} \Big(\widetilde{G}(\tau) + \int_0^\tau \psi_s ds \Big).$$

where $\widetilde{G}(t) = \int_t^\infty (s-t) dG(s)$.

Thus the changepoint detection problem reduces to the optimal stopping problem for ψ_t. In some cases it can be solved explicitly,

Remark: the posterior probability process

In the Bayesian setting, another process is often used:

$$\pi_t = \mathrm{P}^G(\theta \leqslant t \mid \mathcal{F}_t)$$

(assuming θ is defined as a random variable).

It can be shown that

$$\pi_t = \frac{\psi_t}{1 - G(t) + \psi(t)},$$

so the optimal stopping problem above for ψ_t can be equivalently rewritten in terms of π_t.

We will use ψ_t as it will be more convenient in more general problems discussed next.

The SDE for ψ_t

Let $Z_t = \frac{d\mathrm{Q}_t}{d\mathrm{P}_t}$. Then the process ψ_t satisfies the SDE

$$d\psi_t = dG(t) + \frac{\psi_t}{Z_t} dZ_t, \qquad \psi_0 = G(0).$$

Particular convenient cases

1. If the observable process X_t is a diffusion with diffusion coefficient $\sigma(t,x)$ and drift changing from 0 to $\mu(t,x)$, we have

$$d\psi_t = dG(t) + \psi_t \mu(t, X_t) dX_t,$$

so the optimal stopping problem is for a Markov process.

2. If X_t changes only at discrete moments of time $n = 0, 1, 2, \ldots$, so that X_n is a Markov sequence under both P and Q, then with some functions f_k

$$Z_n = \prod_{k=1}^{n} f_k(X_{k-1}, X_k).$$

If the support of $G(t)$ is $\mathbb{N} \cup \{0\}$, then the sequence ψ_n satisfies the following recurrent formula

$$\psi_n = f_n(X_{n-1}, X_n)(\Delta G(n) + \psi_{n-1}).$$

In particular, if X_n is a sequence of independent r.v. under P and Q, then f_k do not depend on X_{k-1} and ψ_n is itself Markov under P, Q.

Examples: Brownian motion

Suppose observed is the process

$$X_t = \mu(t-\theta)^+ + B_t,$$

where B_t is a Brownian motion, θ is a non-negative random variable independent of B_t with a known prior distribution $G(t)$.

We want to find a stopping time τ (with respect to to the filtration of X_t) which minimizes

$$\mathrm{E}^G|\tau - \theta|.$$

In the case when G is the exponential distribution, the solution is well-known. We show how it can be solved for general G.

In this case ψ_t satisfies the SDE (with X_t being a B.M. under P)

$$d\psi_t = dG(t) + \mu\psi_t dX_t, \qquad \psi_0 = G(0).$$

The changepoint detection problem reduces to the following standard Markovian optimal stopping problem

$$\min_{\tau} \mathrm{E}^{\mathrm{P}}\Big(\widetilde{G}(\tau) + \int_0^{\tau} \psi_s ds\Big).$$

General methods of solutions of these problems can be found in Peskir & Shiryaev (2006).

Proposition 2. Under some conditions on $G(t)$, the optimal stopping time is

$$\tau^* = \inf\{t \geqslant 0 : \psi_t \geqslant a(t)\},$$

where the stopping boundary $a(t)$ is the unique solution of the integral equation

$$\int_t^T \mathrm{E}^{\mathrm{P}}_{t,a(t)}(\psi_s + \widetilde{G}'(s))\mathbf{I}((\psi_s < a(s))ds = 0$$

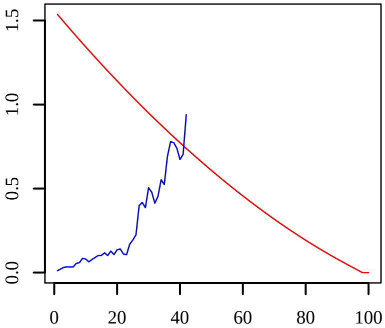
satisfying the conditions

$$a(t) \geqslant -\widetilde{G}'(t), \qquad a(T) = 0.$$

where $[0, T]$ is the support of $G(t)$.

The integral equation is typically solved by discretizing the time and finding $a(t_i)$ backwards.

An illustration for the case of uniform distribution $G(t)$:



Remark: exponential distribution $G(t)$

When $G(t)$ is exponential, the stopping boundary is constant for the process π_t:

$$\tau^* = \inf\{t \geqslant 0 : \pi_t \geqslant a\},$$

where a can be found explicitly as a root of some algebraic equation (Shiryaev, 1963).

Generalizations

The above problem and its solution can be generalized as follows.

1. It is possible to consider a wider class of penalty functions, which assigns linear or exponential penalty for detection delay:

$$\min_{\tau} \int_0^{\infty} \mathrm{E}^{\theta} H(\tau - \theta)\, dG(\theta),$$

where the penalty function

$$H(t) = \begin{cases} \text{non-increasing}, & t \leqslant 0 \\ 0, & t = 0 \\ ct \text{ or } c(e^{bt} - 1), & t \geqslant 0 \end{cases}$$

In this case the changepoint detection problem reduces to the optimal stopping problem

$$\min_{\tau} \mathrm{E}^P\Big(\widetilde{H}(\tau) + \int_0^\tau \psi_s^{(b)} ds\Big),$$

where

$$\widetilde{H}(t) = \int_t^\infty H(t-s)dG(s)$$

and $\psi_t^{(b)}$ is the generalized Shiryaev–Roberts statistic, which satisfies the SDE

$$d\psi_t^{(b)} = b\psi_t^{(b)}dt + dG(t) + \frac{\psi_t^{(b)}}{Z_t}dZ_t, \qquad \psi_y^{(b)} = G(0).$$

2. It is possible to consider a conditional optimality criterion: given two penalty functions $H_1(t)$ and $H_2(t)$ as above,

$$\text{minimize } \int_0^\infty \mathrm{E}^\theta H_1(\tau-\theta)dG(\theta)$$

over stopping times τ satisfying the following condition for some α:

$$\int_0^\infty \mathrm{E}^\theta H_2(\tau-\theta)dG(\theta) \leqslant \alpha.$$

For example if $H_1(t) = \max(t, 0)$ and $H_2(t) = \mathbf{I}(t < 0)$ we get the classical problem of minimizing the detection delay given given the maximum probability of a false alarm.

This setting reduces to the above one by Lagrange multipliers.

Confidence intervals for θ

Choosing $H(t) = \mathbf{I}(|t| > \varepsilon)$, the above problem becomes the problem of finding the best online confidence interval of length 2ε:

$$\text{maximize } \mathrm{P}^G(|\tau - \theta| \leqslant \varepsilon).$$

This problem is interesting as it does not reduce to a Markov optimal stopping problem.

Proposition 3. We have

$$\max_\tau \mathrm{P}^G(|\tau - \theta| \leqslant \varepsilon) = \max_\tau \mathrm{E}^{\mathrm{P}}\big(\psi_\tau - \psi_{\tau-\varepsilon} + G(\tau + h) - G(\tau)\big).$$

($\tau - \varepsilon$ is not a stopping time)

Although it is not known how to solve such an optimal problem analytically, we can obtain, in some sense, an optimal stopping time τ^* when $\varepsilon \to 0$, which turns out to be deterministic.

Namely, for a stopping time τ define

$$\mathcal{R}(\tau) = \lim_{\varepsilon \to 0} \frac{\mathrm{P}^G(|\tau - \theta| \leqslant \varepsilon)}{2\varepsilon}.$$

Proposition 4. Suppose θ takes on values in a bounded interval $[0, T]$ and $G(t)$ has a continuous density on $[0, T]$, which attains its maximum at a point $t^* \in (0, T)$.

Then $\tau^* = t^*$ and $\mathcal{R}(\tau^*) = g(t^*)$.

Optimal stopping problems with changepoints

Let $(S_t)_{t \geqslant 0}$ be a geometric Brownian motion with a changepoint: a process such that

$$\frac{dS_t}{S_t} = \begin{cases} \mu_1 dt + \sigma dB_t, & t < \theta, \\ \mu_2 dt + \sigma dB_t, & t \geqslant \theta, \end{cases} \qquad S_0 = 1,$$

where

- $B = (B_t)_{t \geqslant 0}$ is a standard Brownian motion;
- $\sigma > 0$ and $\mu_1 > 0 > \mu_2$ are known parameters;
- θ is the changepoint uniformly distributed on $(0, 1)$, i.e. $G(t) = G(0) + \rho t$ for $t \in [0, T)$, where $\rho \leqslant 1$.

We consider the following optimal stopping problem:

$$V = \sup_{\tau \leqslant 1} \mathrm{E} S_\tau.$$

This problem was proposed by Beibel & Lerche (1997), and later also studied by Novikov & Shiryaev (2008), Ekstrom & Lindberg (2012) and others when θ is exponentially distributed.

By changing the parameters the result can be extended to the problems $V_\alpha = \sup_{\tau \leqslant 1} \mathrm{E} S_\tau^\alpha$ and also $V_0 = \sup_{\tau \leqslant 1} \mathrm{E} \log(S_\tau)$.

A solution of the problem

Let $X_t = (\log S_t - \mu_1 t - \sigma^2 t/2)/\sigma$, $\quad \mu = (\mu_1 - \mu_2)/\sigma$.

Introduce the new measure $\widetilde{\mathrm{P}}$ such that

$(X_t - \sigma t)$ is a $\widetilde{\mathrm{P}}$ – Brownian motion.

We find that for any stopping time $\tau \leqslant 1$

$$\mathrm{E}^G S_\tau = \mathrm{E}^{\widetilde{\mathrm{P}}}\Big[e^{\mu_1 \tau}(\psi_\tau + 1 - \tau)\Big],$$

where the statistic ψ_t here is given by

$$\psi_t = e^{-\mu X_t - \mu^2 t/2}\left(G(0) + \int_0^t e^{\mu X_s + \mu^2 s/2} \rho ds\right).$$

Proposition 5. The optimal stopping time is

$$\tau^* = \inf\{t \geqslant 0 : \psi_t \geqslant a^*(t)\},$$

where $a^*(t)$ is the unique continuous solution of the equation

$$\int_t^1 e^{\mu_1 s} \mathrm{E}^{\widetilde{\mathrm{P}}}\Big[(\mu_2 \psi_s + \mu_1(1-s))\mathbf{I}(\psi_s \leqslant a^*(s)) \,\Big|\, \psi_t = a^*(t)\Big] ds = 0,$$

satisfying the conditions

$$a^*(t) \geqslant \frac{\mu_1}{|\mu_2|}(1 - G(0) - \rho t) \text{ for } t < 1, \qquad a^*(1) = \frac{\mu_1}{|\mu_2|}(1 - G(0) - \rho).$$

The value $V = \mathrm{E} S_{\tau^*}$ can be found from the formula

$$V = \int_0^1 e^{\mu_1 s} \mathrm{E}^{\widetilde{\mathrm{P}}}[\mu_2 \psi_s + \mu_1(1-s)]\mathbf{I}(\psi_s < a^*(s)) ds + 1.$$

A discrete-time version

In the discrete time it makes sense to consider the problem when also the volatility can change.

Suppose we sequentially observe a random sequence $\widetilde{S}_0, \widetilde{S}_1, \ldots, \widetilde{S}_N$ of the following structure

$$\log \frac{\widetilde{S}_n}{\widetilde{S}_{n-1}} = \begin{cases} \mu_1 + \sigma_1 \xi_n, & n < \theta \\ \mu_2 + \sigma_2 \xi_n, & n \geqslant \theta \end{cases}, \qquad \widetilde{S}_0 = 1.$$

where μ, σ are known parameters, ξ_i are i.i.d. $\mathcal{N}(0,1)$ random variables, θ is the changepoint with prior probabilities $p_k = \mathrm{P}^G(\tau = k)$.

We consider the following discrete-time optimal stopping problem:

$$\widetilde{V} = \sup_{\tau \leqslant N} \mathrm{E}\widetilde{S}_\tau.$$

The Shiryaev–Roberts statistic here is given by

$$\widetilde{\psi}_n = (p_n + \widetilde{\psi}_{n-1}) \cdot \frac{\sigma^1}{\sigma^2} \exp\Big(\frac{(\widetilde{X}_n - \mu_1)^2}{2\sigma_1^2} - \frac{(\widetilde{X}_n - \mu_2)^2}{2\sigma_2^2} \Big),$$

where $\widetilde{X}_n = \log(\widetilde{S}_n / \widetilde{S}_{n-1})$.

Proposition 6. The optimal stopping time is given by

$$\widetilde{\tau}^* = \inf\{0 \leqslant n \leqslant N : \widetilde{\psi}_n \geqslant \widetilde{a}^*(n)\},$$

where

$$\widetilde{a}^*(n) = \inf\{x \geqslant 0 : \widetilde{V}_n(x) = 0\}$$

for the family of functions $\widetilde{V}_0, \widetilde{V}_1, \ldots, \widetilde{V}_N$, which are non-decreasing and can be found from some backward induction formula starting from $\widetilde{V}_N$.

Applications

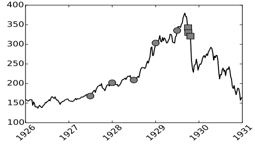
On several financial time series, we show how the above solution can be applied and compare different choices of parameters μ, σ.

We take sequences of prices where changes of trends in the whole sequences can be seen by eye, but we want to detect them from sequential observations.

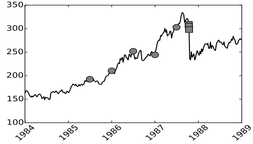
Altogether 9 cases: the US stock market in 1929, 1987, 2001, 2008 and 2015; the Japanese stock and land markets in 1990; Iceland in 2008; the Chinese market in 2015; Apple stock in 2012,

(Economic discussions can be found in e.g. Lleo & Ziemba (2012, 2015)

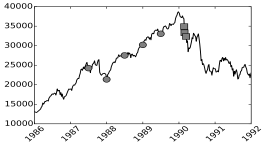
A: US 1929



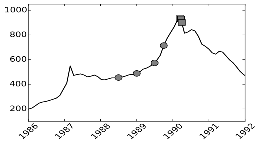
B: US 1987



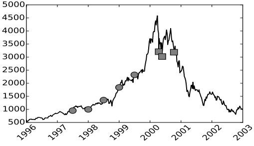
C: Japan 1990



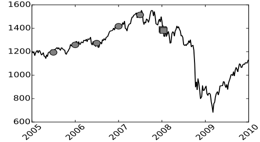
D: Japan land 1990



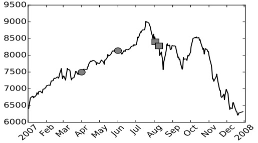
E: US 2001



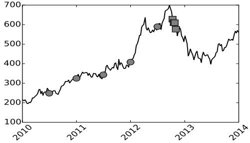
F: US 2008



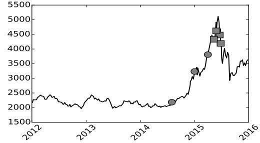
G: Iceland 2008



H: Apple 2012



I: China 2015



Assumptions

1. We observe a sequence of stock prices (or index values) $S_0, S_1, \ldots, S_T$, which has a positive trend initially.
2. It is assumed that the stock price follows the model

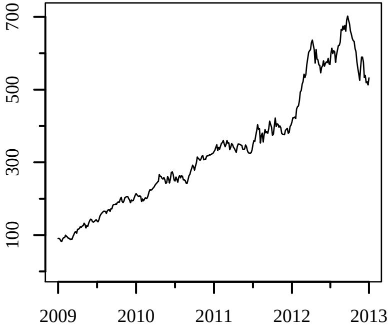
$$\log \frac{S_t}{S_{t-1}} = \begin{cases} \mu_1 + \sigma_1 \xi_t, & t < \theta, \\ \mu_2 + \sigma_2 \xi_t, & t \geqslant \theta, \end{cases}$$

   where $\theta \in \{1, \ldots, T+1\}$ is a uniform random variable.
3. The parameters μ_1, σ_1 are estimated using the previous data $S_{-t_0}, \ldots, S_{-1}, S_0$.

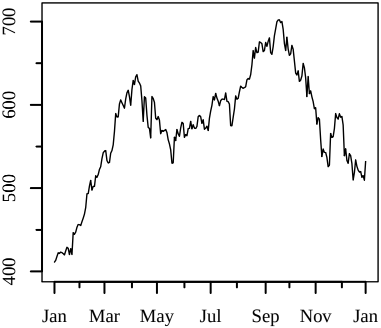
   The choice of the vallues of μ_2, σ_2 is subjective and we compare different possibilities.

Example 1: Apple Inc, 2009-2012

During 2009-2012, Apple's stock price increased almost 9 times, from \$82.33 (6-Mar-09), to \$705.07 (21-Sep-12). By the end of 2012 it fell to \$532.17.



Apple's stock in 2009-2012



Apple's stock in 2012 .

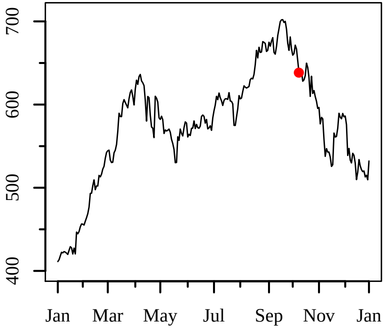
We assume $T \sim$ 31 Dec. 2012.

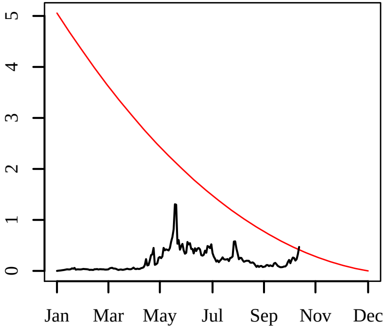
Buy	**Sell**	**% of max.**
3-Jan-11 ($ 329.57)	9-Oct-12 ($ 635.85)	90.56
1-Jul-11 ($ 343.26)	8-Oct-12 ($ 638.17)	90.89
3-Jan-12 ($ 411.23)	8-Oct-12 ($ 638.17)	90.89
1-May-12 ($ 582.13)	9-Oct-12 ($ 635.85)	90.56
3-Jul-12 ($ 599.41)	9-Oct-12 ($ 635.85)	90.56
1-Aug-12 ($ 606.81)	11-Oct-12 ($ 628.10)	89.46

On the graphs – the result of applying the method starting on January 3, 2012.

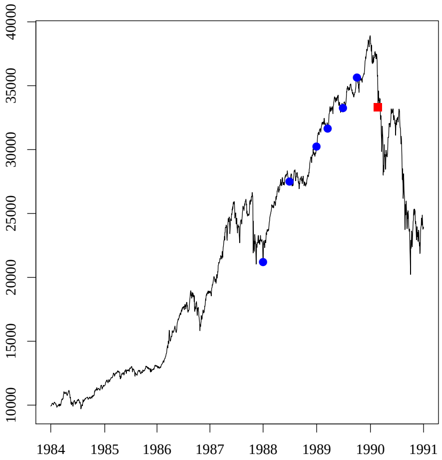
Left – the graph of the price (the red point is the selling price).

Right – the statistic ψ and the optimal stopping boundary.



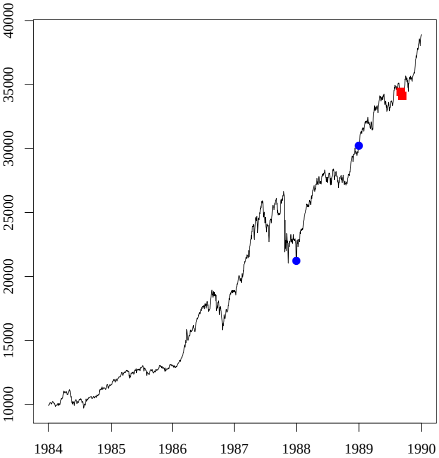


Example 2: NIKKEI-225 in the late 1980s We applied the model for starting dates between 1988 and 1990, which gave the exit date 90-02-26. The assumption is $T \sim$ the end of 1990.



Buy	**Sell**	**% of max.**
88-01-04 (21217.00)	90-02-26 (33322.00)	85.63
88-07-01 (27504.00)	90-02-26 (33322.00)	85.63
89-01-04 (30244.00)	90-02-26 (33322.00)	85.63
89-04-03 (33042.00)	90-02-26 (33322.00)	85.63
89-07-03 (33236.00)	90-02-26 (33322.00)	85.63
89-10-02 (35623.00)	90-02-26 (33322.00)	85.63

A good exit point is also obtained if the disorder does not happen: below is the graph when $T \sim$ the end of 1989.



The results when the disorder does not happen are as follows.

Buy	**Sell**	**% of max.**
88-01-04 (21217.00)	89-08-31 (34431.00)	88.48
88-07-01 (27504.00)	89-09-07 (34153.00)	87.76
89-01-04 (30244.00)	89-09-08 (34116.00)	87.67
89-04-03 (33042.00)	89-10-12 (34795.00)	89.41
89-07-03 (33236.00)	89-10-16 (34469.00)	88.57

Comparing different choices of parameters by their scores

	$\mu_2 = -\mu$ $\sigma_1 = \sigma_2$	$\mu_2 = -2\mu$ $\sigma_1 = \sigma_2$	$\mu_2 = -2\mu$ $\sigma_1 = 2\sigma_2$	$\mu_2 = -3\mu$ $\sigma_1 = \sigma_2$	$\mu_2 = -3\mu$ $\sigma_1 = 3\sigma_2$
A	2	5	1	5	1
B	2	4	1	4	1
C	1	4	4	4	4
D	5	7	4	6	0
E	1	4	0	8	2
F	0	4	0	3	4
G	0	1	1	2	0
H	2	6	0	6	2
I	0	2	0	3	1

Results of the parameters in % of the maximum price on average

	$\mu_2 = -2\mu$ $\sigma_1 = \sigma_2$	$\mu_2 = -3\mu$ $\sigma_1 = \sigma_2$
A	0.85	0.82
B	0.87	0.81
C	0.83	0.81
D	0.96	0.97
E	0.77	0.72
F	0.89	0.9
G	0.91	0.93
H	0.81	0.80
I	0.85	0.87

Thank you for your attention