R. S. Liptser
A. N. Shiryayev

Statistics of Random Processes I

General Theory

Translated by A. B. Aries

A17

R. S. Liptser
A. N. Shiryayev

Statistics of Random Processes I

General Theory

Translated by A. B. Aries

Springer-Verlag
New York Heidelberg Berlin

R. S. Liptser
Institute for Problems of Control Theory
Moscow
Profsojuznaja 81
U.S.S.R.

A. N. Shiryayev
Institute of Control Sciences
Moscow 117806
U.S.S.R.

Title of the Russian Original Edition: *Statistika sluchaĭnyk protsessov.* Nauka, Moscow, 1974.

AMS Subject Classifications: 60Gxx, 60Hxx, 60Jxx, 62Lxx, 62Mxx, 62Nxx, 93Exx, 94A05

Library of Congress Cataloging in Publication Data

Liptser, Robert Shevilevich
Statistics of random processes.
(Applications of mathematics; v. 5)
Revised and expanded translation of the authors' Statistika sluchaĭnyk protsessov, originally published in 1974.
Bibliography: p.
1. Stochastic processes 2 Mathematical statistics. I. Shiriaev, Al'bert Nikolaevich, joint author. II. Title.
QA274.S53 519.2 76-49817

Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-90226-0 Springer-Verlag New York
ISBN 3-540-90226-0 Springer-Verlag Berlin Heidelberg

Contents of volume II

Contents

Contents of volume I

Introduction

A considerable number of problems in the statistics of random processes are formulated within the following scheme.

On a certain probability space $(\Omega, \mathscr{F}, P)$ a partially observable random process $(\theta, \xi) = (\theta_t, \xi_t)$, $t \geq 0$, is given with only the second component $\xi = (\xi_t)$, $t \geq 0$, observed. At any time t it is required, based on $\xi_0^t = \{\xi_s, 0 \leq s \leq t\}$, to estimate the unobservable state θ_t. This problem of estimating (in other words, the *filtering* problem) θ_t from ξ_0^t will be discussed in this book.

It is well known that if $M(\theta_t^2) < \infty$, then the optimal mean square estimate of θ_t from ξ_0^t is the a posteriori mean $m_t = M(\theta_t | \mathscr{F}_t^\xi)$, where $\mathscr{F}_t^\xi = \sigma\{\omega: \xi_s, s \leq t\}$ is the σ-algebra generated by ξ_0^t. Therefore, the solution of the problem of optimal (in the mean square sense) filtering is reduced to finding the conditional (mathematical) expectation $m_t = M(\theta_t | \mathscr{F}_t^\xi)$.

In principle, the conditional expectation $M(\theta_t | \mathscr{F}_t^\xi)$ can be computed by Bayes' formula. However, even in many rather simple cases, equations obtained by Bayes' formula are too cumbersome, and present difficulties in their practical application as well as in the investigation of the structure and properties of the solution.

From a computational point of view it is desirable that the formulae defining the *filter* $m_t, t \geq 0$, should be of a recurrent nature. Roughly speaking, it means that $m_{t+\Delta}$, $\Delta > 0$, must be built up from m_t and observations $\xi_t^{t+\Delta} = \{\xi_s : t \leq s \leq t + \Delta\}$. In the discrete case $t = 0, 1, 2, \ldots$, the simplest form of such recurrence relations can be, for example, the equation

$$\Delta m_t = a(t, m_t) + b(t, m_t)(\xi_{t+1} - \xi_t), \tag{1}$$

where $\Delta m_t = m_{t+1} - m_t$. In the case of continuous time, $t \geq 0$, stochastic differential equations

$$dm_t = a(t, m_t)dt + b(t, m_t)d\xi_t \tag{2}$$

have such a form.

It is evident that without special assumptions concerning the structure of the process (θ, ξ) it is difficult to expect that optimal values m_t should satisfy recurrence relations of the types given by (1) and (2). Therefore, before describing the structure of the process (θ, ξ) whose filtering problems are investigated in this book, we shall study a few specific examples.

Let θ be a Gaussian random variable with $M\theta = m$, $D\theta = \gamma$, which for short will be written $\theta \sim N(m, \gamma)$. Assume that the sequence

$$\xi_t = \theta + \varepsilon_t, \qquad t = 1, 2, \ldots, \tag{3}$$

is observed, where $\varepsilon_1, \varepsilon_2, \ldots$ is a sequence of mutually independent Gaussian random variables with zero mean and unit dispersion independent also of θ. Using a theorem on normal correlation (Theorem 13.1) it is easily shown that $m_t = M(\theta | \xi_1, \ldots, \xi_t)$. The *tracking* errors $\gamma_t = M(\theta - m_t)^2$ are found by

$$m_t = \frac{m + \sum_{i=1}^t \xi_i}{1 + \gamma t}, \qquad \gamma_t = \frac{\gamma}{1 + \gamma t}. \tag{4}$$

From this we obtain the following recurrence equations for m_t and γ_t:

$$\Delta m_t = \frac{\gamma_t}{1 + \gamma t}[\xi_{t+1} - m_t], \tag{5}$$

$$\Delta \gamma_t = -\frac{\gamma_t^2}{1 + \gamma_t}, \tag{6}$$

where $\Delta m_t = m_{t+1} - m_t$, $\Delta \gamma_t = \gamma_{t+1} - \gamma_t$.

Let us make this example more complicated. Let θ and $\xi_1, \xi_2, \ldots$ be the same as in the previous example, and let the observable process ξ_t, $t = 1, 2, \ldots$, be defined by the relations

$$\xi_{t+1} = A_0(t, \xi) + A_1(t, \xi)\theta + \varepsilon_{t+1}, \tag{7}$$

where functions $A_0(t, \xi)$ and $A_1(t, \xi)$ are assumed to be $\mathscr{F}_t^\xi$-measurable (i.e., $A_0(t, \xi)$ and $A_1(t, \xi)$ at any time depend only on the values $(\xi_0, \ldots, \xi_t)$), $\mathscr{F}_t^\xi = \sigma\{\omega: \xi_0, \ldots, \xi_t\}$.

The necessity to consider the coefficients $A_0(t, \xi)$ and $A_1(t, \xi)$ for all "past history" values $(\xi_0, \ldots, \xi_1)$ arises, for example, in control problems (Section 14.3), where these coefficients play the role of "controlling" actions, and also in problems of information theory (Section 16.4), where the pair of functions $(A_0(t, \xi), A_1(t, \xi))$, is treated as "coding" using noiseless feedback.

It turns out that for the scheme given by (7) the optimal value $m_t = M(\theta_t | \mathscr{F}_t^\xi)$ and the conditional dispersion $\gamma_t = M[(\theta - m_t)^2 | \mathscr{F}_t^\xi]$ also satisfy recurrence equations (see Section 13.5):

$$\Delta m_t = \frac{\gamma_t A_1(t, \xi)}{1 + A_1^2(t, \xi)\gamma_t}(\xi_{t+1} - A_0(t, \xi) - A_1(t, \xi)m_t), \qquad m_0 = m; \tag{8}$$

$$\Delta \gamma_t = -\frac{A_1^2(t, \xi)\gamma_t^2}{1 + A_1^2(t, \xi)\gamma_t}, \qquad \gamma_0 = \gamma. \tag{9}$$

In the schemes given by (3) and (7) the question, in essence, is a traditional problem of mathematical statistics—Bayes' estimation of a random parameter from the observations ξ_0^t. The next step to make the scheme given by (7) more complicated is to consider a random process θ_t rather than a random variable θ.

Assume that the random process $(\theta, \xi) = (\theta_t, \xi_t)$, $t = 0, 1, \ldots$, is described by the recurrence equations

$$\begin{aligned} \theta_{t+1} &= a_0(t, \xi) + a_1(t, \xi)\theta_t + b(t, \xi)\varepsilon_1(t + 1), \\ \xi_{t+1} &= A_0(t, \xi) + A_1(t, \xi)\theta_t + B(t, \xi)\varepsilon_2(t + 1) \end{aligned} \tag{10}$$

where $\varepsilon_1(t)$, $\varepsilon_2(t)$, $t = 1, 2, \ldots$, the sequence of independent variables, is normally distributed, $N(0, 1)$, and also indépendent of (θ_0, ξ_0). The coefficients $a_0(t, \xi), \ldots, B(t, \xi)$ are assumed to be $\mathscr{F}_t^\xi$-measurable for any $t = 0, 1 \ldots$

In order to obtain recurrence equations for estimating $m_t = M(\theta_t | \mathscr{F}_t^\xi)$ and conditional dispersion $\gamma_t = M\{[\theta_t - m_t]^2 | \mathscr{F}_t^\xi\}$, let us assume that the conditional distribution $P(\theta_0 \le x | \xi_0)$ is (for almost all ξ_0) normal, $N(m, \gamma)$. The essence of this assumption is that it permits us to prove (see Chapter 13) that the sequence (θ, ξ) satisfying (10) is conditionally Gaussian. This means, in particular, that the conditional distribution $P(\theta_t \le x | \mathscr{F}_t^\xi)$ is (almost surely) Gaussian. But such a distribution is characterized only by its two conditional moments m_t and γ_t, leading to the following closed system of equations:

$$\begin{aligned} m_{t+1} &= a_0 + a_1 m_t + \frac{a_1 A_1 \gamma_t}{B^2 + A_1^2 \gamma_t}[\xi_{t+1} - A_0 - A_1 m_t], \qquad m_0 = m; \\ \gamma_{t+1} &= [a_1^2 \gamma_t + b^2] - \frac{(a_1 A_1 \gamma_t)^2}{B^2 + A_1^2 \gamma_t}, \qquad \gamma_0 = \gamma \end{aligned} \tag{11}$$

(in the coefficients $a_0, \ldots, B$, for the sake of simplicity, arguments t and ξ are omitted).

The equations in (11) are deduced (in a somewhat more general framework) in Chapter 13. Their deduction does not need anything except the theorem of normal correlation. In this chapter, equations for optimal estimation in extrapolation problems (estimating θ_τ from ξ_0^t, when $\tau > t$) and interpolation problems (estimating θ_τ from ξ_0^t when $\tau < t$) are derived. Chapter 14 deals with applications of these equations to various statistical problems of random sequences, to control problems, and to problems of constructing pseudosolutions to linear algebraic systems.

These two chapters can be read independently of the rest of the book, and this is where the reader should start if he is interested in nonlinear filtering problems but is not sufficiently acquainted with the general theory of random processes.

The main part of the book concerns problems of optimal filtering and control (and also related problems of interpolation, extrapolation, sequential estimation, testing of hypotheses, etc.) in the case of *continuous* time. These problems are interesting per se; and, in addition, easy formulations and compact formulae can be obtained for them. It should be added that often it is easier, at first, to study the continuous analog of problems formulated for discrete time, and use the results obtained in the solution of the latter.

The simplicity of formulation in the case of continuous time is, however, not easy to achieve—rather complicated techniques of the theory of random processes have to be invoked. Later on, we will discuss the methods and the techniques used in this book in more detail, but here we consider particular cases of the filtering problem for the sake of illustration.

Assume that the partially observable random process $(\theta, \xi) = (\theta_t, \xi_t)$, $t \geq 0$, is Gaussian, governed by stochastic differential equations (compare with the system (10)):

$$d\theta_t = a(t)\theta_t dt + b(t)dw_1(t), \qquad d\xi_t = A(t)\theta_t\,dt + B(t)dw_2(t), \qquad \theta_0 \equiv 0, \tag{12}$$

where $w_1(t)$ and $w_2(t)$ are standard Wiener processes, mutually independent and independent of (θ_0, ξ_0), and $B(t) \geq C > 0$. Let us consider the component $\theta = (\theta_t)$, $t \geq 0$, as unobservable. The filtering problem is that of optimal estimation of θ_t from ξ_0^t in the mean square sense for any $t \geq 0$.

The process (θ, ξ), according to our assumption, is Gaussian; hence the optimal estimate $m_t = M(\theta_t | \mathscr{F}_t^\xi)$ depends linearly on $\xi_0^t = \{\xi_s : s \leq t\}$. More precisely, there exists (Lemma 10.1) a function $G(t, s)$, with $\int_0^t G^2(t, s)ds < \infty$, $t > 0$, such that (almost surely):

$$m_t = \int_0^t G(t, s)d\xi_s. \tag{13}$$

If this expression is formally differentiated, we obtain

$$dm_t = G(t, t)d\xi_t + \left(\int_0^t \frac{\partial G(t, s)}{\partial t}\, d\xi_s\right)dt. \tag{14}$$

The right side of this equation can be transformed using the fact that the function $G(t, s)$ satisfies the Wiener–Hopf equation (see (10.25)), which in our case reduces to

$$\frac{\partial G(t, s)}{\partial t} = \left[a(t) - \gamma_t \frac{A^2(t)}{B^2(t)}\right] G(t, s), \qquad t > s, \tag{15}$$

$$G(s, s) = \frac{\gamma_s A(s)}{B^2(s)}, \qquad \gamma_s = M[\theta_s - m_s]^2. \tag{16}$$

Taking into account (15) and (14), we infer that the optimal estimate $m_t, t > 0$, satisfies a linear stochastic differential equation,

$$dm_t = a(t)m_t\,dt + \frac{\gamma_t A(t)}{B^2(t)}[d\xi_t - A(t)m_t\,dt]. \tag{17}$$

This equation includes the tracking error $\gamma = M[\theta_t - m_t]^2$, which in turn is the solution of the Ricatti equation

$$\dot{\gamma}_t = 2a(t)\gamma_t - \frac{A^2(t)\gamma_t^2}{B^2(t)} + b^2(t). \tag{18}$$

(Equation (18) is easy to obtain applying the Ito formula for substitution of variables to the square of the process $[\theta_t - m_t]$ with posterior averaging.)

Let us discuss Equation (17) in more detail taking, for simplicity, $\xi_0 \equiv 0$. Denote

$$\overline{w} = \int_0^t \frac{d\xi_s - A(s)m_s\,ds}{B(s)}. \tag{19}$$

Then Equation (17) can be rewritten:

$$dm_t = a(t)m_t\,dt + \frac{\gamma_t A(t)}{B(t)}\,d\overline{w}_t. \tag{20}$$

The process $(\overline{w}_t)$, $t \geq 0$, is rather remarkable and plays a key role in filtering problems. The point is that, first, this process turns out to be a Wiener process (with respect to the σ-algebras $(\mathscr{F}_t^\xi)$, $t \geq 0$), and secondly, it contains the same *information* as the process ξ does. More precisely, it means that for all $t \geq 0$, the σ-algebras $\mathscr{F}_t^{\overline{w}} = \sigma\{\omega : \overline{w}_s, s \leq t\}$ and $\mathscr{F}_t^\xi = \sigma\{\omega : \xi_s, s \leq t\}$ coincide:

$$\mathscr{F}_t^{\overline{w}} = \mathscr{F}_t^\xi, \qquad t \geq 0 \tag{21}$$

(see Theorem 7.16). By virtue of these properties of the process, $\overline{w}$ it is referred to as the *innovation* process.

The equivalence of σ-algebras $\mathscr{F}_t^\xi$ and $\mathscr{F}_t^{\overline{w}}$ suggests that for m_t not only is Equation (13) justified but also the representation

$$m_t = \int_0^t F(t, s)\,d\overline{w}_s \tag{22}$$

where $\overline{w} = (\overline{w}_t)$, $t \geq 0$ is the innovation process, and functions $F(t, s)$ are such that $\int_0^t F^2(t, s)\,ds < \infty$. In the main part of the text (Theorem (7.16)) it is shown that the representation given by (22) can actually be obtained from results on the structure of functionals of diffusion type processes. Equation (20) can be deduced in a simpler way from the representation given by (22) than the representation given by (13). It should be noted, however, that the proof of (22) is more difficult than that of (13).

In this example, the optimal (Kalman–Bucy) filter was linear because of the assumption that the process (θ, ξ) is Gaussian. Let us take now an example where the optimal filter is nonlinear.

Let (θ_t), $t \geq 0$, be a Markov process starting at zero with two states 0 and 1 and the only transition $0 \rightarrow 1$ at a random moment σ, distributed (due to assumed Markov behavior) exponentially: $P(\sigma > t) = e^{-\lambda t}$, $\lambda > 0$. Assume that the observable process $\xi = (\xi_t)$, $t \geq 0$, has a differential

$$d\xi_t = \theta_t \, dt + dw_t \qquad \xi_0 = 0, \tag{23}$$

where $w = (w_t)$, $t \geq 0$, is a Wiener process independent of the process $\theta = (\theta_t)$, $t \geq 0$.

We shall interpret the transition of the process θ from the "zero" state into the unit state as *the occurrence of discontinuity* (at the moment σ). There arises the following problem: to determine at any $t > 0$ from observations ξ_0^t whether or not discontinuity has occurred before this moment.

Denote $\pi_t = P(\theta_t = 1 | \mathscr{F}_t^\xi) = P(\sigma \leq t | \mathscr{F}_t^\xi)$. It is evident that $\pi_t = m_t = M(\theta_t | \mathscr{F}_t^\xi)$. Therefore, the a posteriori probability π_t, $t \geq 0$, is the optimal (in the mean square sense) state estimate of an unobservable process $\theta = (\theta_t)$, $t \geq 0$.

For the a posteriori probability π_t, $t \geq 0$, we can deduce (using, for example, Bayes' formula and results with respect to a derivative of the measure corresponding to the process ξ, with respect to the Wiener measure) the following stochastic differential equation:

$$d\pi_t = \lambda(1 - \pi_t)dt + \pi_t(1 - \pi_t)[d\xi_t - \pi_t \, dt], \qquad \pi_0 = 0. \tag{24}$$

It should be emphasized that whereas in the Kalman–Bucy scheme the optimal filter is linear, Equation (24) is essentially nonlinear. Equation (24) defines the *optimal nonlinear filter*.

As in the previous example, the (innovation) process

$$\bar{w}_t = \int_0^t [d\xi_s - \pi_t \, ds], \qquad t \geq 0,$$

turns out to be a Wiener process and $\mathscr{F}_t^{\bar{w}} = \mathscr{F}_t^\xi$, $t \geq 0$. Therefore, Equation (24) can be written in the following equivalent form:

$$d\pi_t = \lambda(1 - \pi_t)dt + \pi_t(1 - \pi_t)d\bar{w}_t, \qquad \pi_0 = 0. \tag{25}$$

It appears that all these examples are within the following general scheme adopted in this book.

Let $(\Omega, \mathscr{F}, P)$ be a certain probability space with a distinguished non-decreasing set of σ-algebras $(\mathscr{F}_t)$, $t \geq 0 (\mathscr{F}_s \subseteq \mathscr{F}_t \subseteq \mathscr{F}, s \leq t)$. In this probability space we are given a partially observable process (θ_t, ξ_t), $t \geq 0$, and an estimated process (h_t), $t \geq 0$, dependent, generally speaking, on both the unobservable process θ_t, $t \geq 0$, and the observable component (ξ_t), $t \geq 0$.

As to the observable process[1] $\xi = (\xi_t, \mathscr{F}_t)$ it will be assumed that it permits a stochastic differential

$$d\xi_t = A_t(\omega)dt + dw_t, \qquad \xi_0 = 0, \tag{26}$$

where $w = (w_t, \mathscr{F}_t)$, $t \geq 0$, is a standard Wiener process (i.e., a square integrable martingale with continuous trajectories with $M[(w_t - w_s)^2 | \mathscr{F}_s] = t - s$, $t \geq s$, and $w_0 = 0$), and $A = (A_t(\omega), \mathscr{F}_t)$, $t \geq 0$, is a certain integrable random process.[2]

The structure of the unobservable process $\theta = (\theta_t, \mathscr{F}_t)$, $t \geq 0$, is not directly concretized, but it is assumed that the estimated process $h = (h_t, \mathscr{F}_t)$, $t \geq 0$, permits the following representation:

$$h_t = h_0 + \int_0^t a_s(\omega)ds + x_t, \qquad t \geq 0, \tag{27}$$

where $a = (a_t(\omega), \mathscr{F}_t)$, $t \geq 0$, is some integrable process, and $x = (x_t, \mathscr{F}_t)$, $t \geq 0$, is a square integrable martingale.

For any integrable process $g = (g_t, \mathscr{F}_t)$, $t \geq 0$, write $\pi_t(g) = M[g_t | \mathscr{F}_t^\xi]$. Then, if $Mg_t^2 < \infty$, $\pi_t(g)$ is the optimal (in the mean square sense) estimate of g_t from $\xi_0^t = \{\xi_s : s \leq t\}$.

One of the main results of this book (Theorem 8.1) states that for $\pi_t(h)$ the following representation is correct:

$$\pi_t(h) = \pi_0(h) + \int_0^t \pi_s(a)ds + \int_0^t \pi_s(D)d\bar{w}_s + \int_0^t [\pi_s(hA) - \pi_s(h)\pi_s(A)]d\bar{w}_s \tag{28}$$

Here $\bar{w} = (\bar{w}_t, \mathscr{F}_t^\xi)$, $t \geq 0$, is a Wiener process (compare with the innovation processes in the two previous examples), and the process $D = (D_t, \mathscr{F}_t)$, $t \geq 0$, characterizes *correlation* between the Wiener process $w = (w_t, \mathscr{F}_t)$, $t \geq 0$, and the martingale $x = (x_t, \mathscr{F}_t)$, $t \geq 0$. More precisely, the process

$$D_t = \frac{d\langle x, w\rangle_t}{dt}, \qquad t \geq 0, \tag{29}$$

where $\langle x, w\rangle_t$ is a random process involved in Doob–Meyer decomposition of the product of the martingales x and w:

$$M[x_t w_t - x_s w_s | \mathscr{F}_s] = M[\langle x, w\rangle_t - \langle x, w\rangle_s | \mathscr{F}_s] \tag{30}$$

We call the representation in (28) the *main equation* of (optimal nonlinear) *filtering*. Most of known results (within the frame of the assumptions given by (26) and (27)) can be deduced from this equation.

[1] $\xi = (\xi_t, \mathscr{F}_t)$ suggests that values ξ_t are $\mathscr{F}_t$-measurable for any $t \geq 0$.

[2] Actually, this book examines processes ξ of a somewhat more general kind (see Chapter 8).

Let us show, for example, in what way the filtering Equations (17) and (18) in the Kalman–Bucy scheme are deduced from (28), taking, for simplicity, $b(t) \equiv B(t) \equiv 1$.

Comparing (12) with (26) and (27), we see that $A_t(\omega) = A(t)\theta_t$, $w_t = w_2(t)$. Assume $h_t = \theta_t$. Then, due to (12),

$$h_t = h_0 + \int_0^t a(s)\theta_s\, ds + w_1(t). \tag{31}$$

The processes $w_1 = (w_1(t))$ and $w_2 = (w_2(t))$, $t \geq 0$, are independent square integrable martingales, hence for them $D_t \equiv 0$(P-a.s.). Then, due to (28), $\pi_t(\theta)$ has a differential

$$d\pi_t(\theta) = a(t)\pi_t(\theta)dt + A(t)[\pi_t(\theta^2) - \pi_t^2(\theta)]d\bar{w}_t, \tag{32}$$

i.e.,

$$dm_t = a(t)m_t\, dt + A(t)\gamma_t\, d\bar{w}_t \tag{33}$$

where we have taken advantage of the Gaussian behavior of the process (θ, ξ); (P-a.s.)

$$\pi_t(\theta^2) - \pi_t^2(\theta) = M[(\theta_t - m_t)^2 | \mathscr{F}_t^\xi] = M[\theta_t - m_t]^2 = \gamma_t.$$

In order to deduce an equation for γ_t from (28), we take $h_t = \theta_t^2$. Then, from the first equation of the system given by (12) we obtain by the Ito formula for substitution of variables (Theorem 4.4),

$$\theta_t^2 = \theta_0^2 + \int_0^t a_s(\omega)ds + x_t, \tag{34}$$

where

$$a_s(\omega) = 2a(s)\theta_s^2 + b^2(s)$$

and

$$x_t = \int_0^t \theta_s\, dw_1(s).$$

Therefore, according to (28)

$$d\pi_t(\theta^2) = [2a(t)\pi_t(\theta^2) + b^2(s)(t)]dt + A(t)[\pi_t(\theta^3) - \pi_t(\theta)\pi_t(\theta^2)]d\bar{w}_t. \tag{35}$$

From (32) and (35) it is seen that in using the main filtering equation, (28), we face the difficulty that for finding conditional lower moments a knowledge of higher moments is required. Thus, for finding equations for $\pi_t(\theta^2)$ the knowledge of the third a posteriori moment $\pi_t(\theta^3) = M(\theta_t^3 | \mathscr{F}_t^\xi)$ is required. In the case considered this difficulty is easy to overcome, since, due to the Gaussian behavior of the process (θ, ξ), the moments $\pi_t(\theta^n) = M(\theta_t^n | \mathscr{F}_t^\xi)$ for all $n \geq 3$ are expressed through $\pi_t(\theta)$ and $\pi_t(\theta^2)$. In particular, $\pi_t(\theta^3) - \pi_t(\theta)\pi_t(\theta^2) = M[\theta_t^2(\theta_t - m_t) | \mathscr{F}_t^\xi] = 2m_t\gamma_t$ and, therefore,

$$d\pi_t(\theta^2) = [2a(t)\pi_t(\theta^2) + b^2(t)]dt + 2A(t)m_t\gamma_t\, d\bar{w}_t. \tag{36}$$