

June 8, 2012

## SECOND-ORDER CONTINUOUS-TIME NON-STATIONARY GAUSSIAN AUTOREGRESSION

N. LIN AND S. V. LOTOTSKY

ABSTRACT. The objective of the paper is to identify and investigate all possible types of asymptotic behavior for the maximum likelihood estimators of the unknown parameters in the second-order linear stochastic ordinary differential equation driven by Gaussian white noise. The emphasis is on the non-ergodic case, when the roots of the corresponding characteristic equation are not both in the left half-plane.

### 1. INTRODUCTION

Consider the stochastic ordinary differential equation

$$(1.1) \quad \ddot{X}(t) = \theta_1 \dot{X}(t) + \theta_2 X(t) + \sigma \dot{W}(t), \quad t > 0,$$

with a standard Brownian motion  $W = W(t)$ , non-random initial conditions  $X(0)$ ,  $\dot{X}(0)$ , and two real parameters  $\theta_1, \theta_2$ . Equation (1.1) is often referred to as a continuous time auto-regression of second order, or CAR(2), being a particular case of CAR(N)

$$(1.2) \quad X^{(N)} = \sum_{k=0}^{N-1} \theta_{N-k} X^{(k)} + \dot{W};$$

see [4]. A rigorous interpretation of equation (1.1) is the system

$$(1.3) \quad dX = \dot{X}dt, \quad d\dot{X} = (\theta_2 X + \theta_1 \dot{X})dt + \sigma dW(t).$$

In the matrix-vector form, system (1.3) becomes a particular case of the multi-dimensional Ornstein-Uhlenbeck process studied in [2]:

$$(1.4) \quad d\mathbf{X}(t) = \Theta \mathbf{X}(t)dt + \boldsymbol{\sigma} dW(t),$$

with

$$\mathbf{X}(t) = \begin{pmatrix} X(t) \\ \dot{X}(t) \end{pmatrix}, \quad \Theta = \begin{pmatrix} 0 & 1 \\ \theta_2 & \theta_1 \end{pmatrix}, \quad \boldsymbol{\sigma} = \begin{pmatrix} 0 \\ \sigma \end{pmatrix}.$$

The estimator studied in [2] is

$$(1.5) \quad \hat{\Theta}_T = \left( \int_0^T (d\mathbf{X}(t) \mathbf{X}^\top(t)) \right) \left( \int_0^T \mathbf{X}(t) \mathbf{X}^\top(t) dt \right)^{-1}.$$

It is strongly consistent as  $T \rightarrow \infty$ , and this implies strong consistency of the maximum likelihood estimators for  $\theta_1$  and  $\theta_2$  in (1.1) for all  $(\theta_1, \theta_2) \in \mathbb{R}^2$ ; see [2] or Theorem 3.1 below.

When the process  $\mathbf{X}$  defined by (1.4) is ergodic (equivalently, when all eigenvalues of the matrix  $\Theta$  are in the left half-plane), it is known [1, Theorem 4.6.2] that the

---

2000 *Mathematics Subject Classification.* Primary 62F12; Secondary 62F03, 62M07, 62M09.

*Key words and phrases.* Lyapunov Exponent, Maximum Likelihood Estimation, Asymptotic Mixed Normality, Non-Normal Limit Distribution, Rate of Convergence, Second-Order Stochastic Equation.

estimator  $\widehat{\Theta}_T$  is asymptotically normal with rate  $T^{1/2}$ . Some non-ergodic models of the type (1.4) have also been studied (see, for example, [16]). The key object in the analysis of a maximum likelihood estimator (MLE) is the normalized log-likelihood ratio. A desirable property of this normalized log-likelihood ratio is local asymptotic normality (LAN), as it implies certain efficiency of the MLE; see [7, Chapter II]. While the LAN property is typically associated with ergodic models, it can also hold in some non-ergodic models [9]. Still, in many non-ergodic models, LAN is replaced with LAMN (local asymptotic mixed normality), leading to a different kind of efficiency of the MLE; see [12, Chapter 5].

So far, analysis of (1.4) in general and (1.1) in particular was aimed at identifying the particular cases that fit in either LAN or LAMN framework. It turns out that many of the non-ergodic regimes of (1.1) lead to new asymptotic forms of the normalized log-likelihood ratio and to new types of asymptotic behavior of the MLE. In the current paper, we present a complete asymptotic analysis of the MLE and the likelihood ratio for all possible values of the unknown parameters  $(\theta_1, \theta_2) \in \mathbb{R}^2$ . Beside purely theoretical interest, this analysis can help in the investigation of the corresponding discrete-time models (see [4] in the ergodic case).

The rest of the paper is organized as follows. Section 3 states the main results and discusses the results in the broader context of statistical estimation. After some preliminary work in Section 4, the proofs are in Section 5, followed by a brief summary in Section 6.

We fix a stochastic basis  $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$  with a standard Brownian motion  $W = W(t)$ . Other common notations are  $\mathbb{E}$  for the expectation with respect to  $\mathbb{P}$ ,  $\sqrt{-1}$  for the imaginary unit,  $^\top$  for the transpose of a vector or a matrix,  $\dot{g}(t)$ ,  $\ddot{g}(t)$  for the first and second time derivatives of the function  $g$ ,  $\stackrel{d}{=}$  for equality in distribution of random variables,  $\stackrel{\mathcal{L}}{=}$  for equality in law of random processes, and  $0_{\text{a.s.}}(T)$  to denote a process converging to zero with probability one as  $T \rightarrow \infty$ .

## 2. SUMMARY OF THE MAIN RESULTS

To write the likelihood ratio for equation (1.1), define the vectors

$$(2.1) \quad \boldsymbol{\theta} = \begin{pmatrix} \theta_2 \\ \theta_1 \end{pmatrix}, \quad \mathbf{X}(t) = \begin{pmatrix} X(t) \\ \dot{X}(t) \end{pmatrix}.$$

Then equation (1.1) becomes

$$(2.2) \quad d\dot{X} = \boldsymbol{\theta}^\top \mathbf{X} dt + \sigma dW(t).$$

Since  $X(t) = X(0) + \int_0^t \dot{X}(s) ds$ , it follows from (2.2) that

- (1)  $\dot{X} = \dot{X}(t)$  is a diffusion-type process in the sense of Liptser and Shiryaev ([15, Definition 4.2.7]) and generates the measure  $P_T^\boldsymbol{\theta}$  on the space of continuous functions on  $[0, T]$ ,

- (2) For every  $\theta \in \mathbb{R}^2$ , the measure  $P_T^\theta$  is absolutely continuous with respect to the measure  $P_T^0$  generated by the process  $\dot{X}(0) + \sigma W(t)$ ,  $0 \leq t \leq T$ , and

$$(2.3) \quad \frac{dP_T^\theta}{dP_T^0}(\dot{X}) = \exp \left( \frac{1}{\sigma^2} \int_0^T (\theta_2 X(t) + \theta_1 \dot{X}(t)) d\dot{X}(t) - \frac{1}{2\sigma^2} \int_0^T (\theta_2 X(t) + \theta_1 \dot{X}(t))^2 dt \right);$$

see [15, Theorem 7.6].

The expressions for the maximum likelihood estimators  $\hat{\theta}_{1,T}$  of  $\theta_1$  and  $\hat{\theta}_{2,T}$  of  $\theta_2$ , using continuous-time observations of both  $X(t)$ , and  $\dot{X}(t)$ ,  $0 \leq t \leq T$ , now follow from (2.3):

$$(2.4) \quad \begin{aligned} \hat{\theta}_{1,T} &= \frac{\int_0^T X^2(t) dt \int_0^T \dot{X}(t) d\dot{X}(t) - \int_0^T X(t) \dot{X}(t) dt \int_0^T X(t) d\dot{X}(t)}{\int_0^T X^2(t) dt \int_0^T \dot{X}^2(t) dt - \left( \int_0^T X(t) \dot{X}(t) dt \right)^2}, \\ \hat{\theta}_{2,T} &= \frac{\int_0^T \dot{X}^2(t) dt \int_0^T X(t) d\dot{X}(t) - \int_0^T X(t) \dot{X}(t) dt \int_0^T \dot{X}(t) d\dot{X}(t)}{\int_0^T X^2(t) dt \int_0^T \dot{X}^2(t) dt - \left( \int_0^T X(t) \dot{X}(t) dt \right)^2}. \end{aligned}$$

Similar to (2.1) we write

$$(2.5) \quad \hat{\theta}_T = \begin{pmatrix} \hat{\theta}_{2,T} \\ \hat{\theta}_{1,T} \end{pmatrix}.$$

The amount of integration in (2.4) can be reduced using rules of the usual and stochastic calculus and keeping in mind that the processes  $X$  is continuously differentiable and the process  $\dot{X}$  is a continuous semi-martingale with quadratic variation equal to  $\sigma^2 t$ :

$$(2.6) \quad \begin{aligned} \int_0^T X(t) \dot{X}(t) dt &= \int_0^T X(t) dX(t) = \frac{X^2(T) - X^2(0)}{2}, \\ \int_0^T \dot{X}(t) d\dot{X}(t) &= \frac{\dot{X}^2(T) - \sigma^2 T - \dot{X}^2(0)}{2}, \\ \int_0^T X(t) d\dot{X}(t) &= X(T) \dot{X}(T) - X(0) \dot{X}(0) - \int_0^T \dot{X}^2(t) dt. \end{aligned}$$

With the continuous time observations, the value of  $\sigma$  can be assumed known because the quadratic variation process of  $\dot{X}$  at time  $t$ , an observable quantity, is  $\sigma^2 t$ . Note also that  $\sigma$  does not appear in the formulas (2.4).

The measure  $P_T^0$  in (2.3) corresponds to the solution of (2.2) with  $\theta_1 = \theta_2 = 0$ . Sometimes it is more convenient to work with the reference measure coming from some other solution of (1.1), for example, the actual observations. Accordingly, let us fix  $\theta \in \mathbb{R}^2$  and the corresponding observation process  $\mathbf{X} = \mathbf{X}(t)$ ,  $0 \leq t \leq T$ , satisfying

(1.3). Then, for every  $\vartheta \in \mathbb{R}^2$ ,

$$(2.7) \quad \begin{aligned} \ln \frac{dP_T^\vartheta}{dP_T^\theta}(\dot{X}) &= \frac{1}{\sigma^2} \int_0^T (\vartheta^\top \mathbf{X}(t) - \theta^\top \mathbf{X}(t)) d\dot{X}(t) \\ &\quad - \frac{1}{2\sigma^2} \int_0^T \left( (\vartheta^\top \mathbf{X}(t))^2 - (\theta^\top \mathbf{X}(t))^2 \right) dt; \end{aligned}$$

see [15, Theorem 7.19]. With  $\theta$  and  $\mathbf{X}$  fixed, we write

$$L_T(\vartheta) = \ln \frac{dP_T^\vartheta}{dP_T^\theta}(\dot{X}).$$

Define the matrix

$$(2.8) \quad \Psi_T = \int_0^T \mathbf{X}(t) \mathbf{X}^\top(t) dt = \begin{pmatrix} \int_0^T X^2(t) dt & \int_0^T X(t) \dot{X}(t) dt \\ \int_0^T X(t) \dot{X}(t) dt & \int_0^T \dot{X}^2(t) dt \end{pmatrix},$$

Then (2.2) and (2.7) imply

$$(2.9) \quad \begin{aligned} L_T(\vartheta) &= \frac{1}{\sigma} \int_0^T (\vartheta - \theta)^\top \mathbf{X}(t) dW(t) - \frac{1}{2\sigma^2} \int_0^T (\vartheta - \theta)^\top \mathbf{X}(t) \mathbf{X}^\top(t) (\vartheta - \theta) dt \\ &= \frac{1}{\sigma} \int_0^T (\vartheta - \theta)^\top \mathbf{X}(t) dW(t) - \frac{1}{2\sigma^2} (\vartheta - \theta)^\top \Psi(T) (\vartheta - \theta). \end{aligned}$$

We address the following questions about the estimators  $\widehat{\theta}_1(T), \widehat{\theta}_2(T)$ :

- (1) **rate of convergence**, that is, finding positive deterministic functions  $v_i(T)$ ,  $i = 1, 2$ , such that, as  $T \rightarrow \infty$ ,  $v_i(T) \nearrow +\infty$  and  $v_i(T)(\widehat{\theta}_{i,T} - \theta_i)$  converge in distribution to non-degenerate random variables, and identifying the corresponding limit distributions;
- (2) existence of a **normal limit with a random rate** (NLRR), that is, finding a random matrix  $R = R_T$  such that, as  $T \rightarrow \infty$ ,  $R_T(\widehat{\theta}_T - \theta)$  converges in distribution to a Gaussian random vector. The hope is that at least one of the two things happens: (a) the random rate leads to a normal limit when a deterministic rate does not; (b) neither the matrix  $R_T$  nor the parameters of the limit distribution depend explicitly on the  $\theta_1, \theta_2$ .
- (3) **local asymptotic structure** of the normalized log-likelihood ratio, that is, analyzing

$$(2.10) \quad \ell_T(\mathbf{u}) = L_T(\theta + A_T \mathbf{u}),$$

as  $T \rightarrow \infty$ , where  $\theta$  is fixed,  $A_T \in \mathbb{R}^{2 \times 2}$  is a suitably chosen deterministic matrix with  $\lim_{T \rightarrow \infty} |A_T| = 0$  (any matrix norm will work), and  $\mathbf{u} \in \mathbb{R}^2$ .

Let us recall some definitions related to the normalized log-likelihood ratio (2.10). It follows from (2.9) that

$$(2.11) \quad \ell_T(\mathbf{u}) = \frac{1}{\sigma} \int_0^T \left( \mathbf{u}^\top A_T \mathbf{X}(t) \right) dW(t) - \frac{1}{2\sigma^2} \mathbf{u}^\top A_T^\top \Psi_T A_T \mathbf{u}.$$

With a suitable choice of the matrix  $A_T$ , there exists a non-trivial limit in distribution

$$(2.12) \quad \ell_\infty(\mathbf{u}) = \lim_{T \rightarrow \infty} \ell_T(\mathbf{u}).$$

Three particular cases of  $\ell_T$  satisfying (2.12) are of special interest:

- **Local Asymptotic Normality (LAN)**, if there exists a bivariate normal vector  $\boldsymbol{\xi}$  with mean zero and non-degenerate covariance matrix  $\Sigma_{\boldsymbol{\xi}}$  such that, for every  $\mathbf{u} \in \mathbb{R}^2$ ,

$$(2.13) \quad \ell_{\infty}(\mathbf{u}) = \frac{1}{\sigma} \mathbf{u}^{\top} \boldsymbol{\xi} - \frac{1}{2\sigma^2} \mathbf{u}^{\top} \Sigma_{\boldsymbol{\xi}} \mathbf{u}.$$

- **Local Asymptotic Mixed Normality (LAMN)** if there exist a bivariate normal vector  $\boldsymbol{\eta}$  with zero mean and unit covariance matrix, and a random symmetric positive definite matrix  $B \in \mathbb{R}^{2 \times 2}$  such that  $B$  and  $\boldsymbol{\eta}$  are independent and, for every  $\mathbf{u} \in \mathbb{R}^2$ ,

$$(2.14) \quad \ell_{\infty}(\mathbf{u}) = \frac{1}{\sigma} \mathbf{u}^{\top} B^{1/2} \boldsymbol{\eta} - \frac{1}{2\sigma^2} \mathbf{u}^{\top} B \mathbf{u}.$$

If (2.14) holds with a degenerate matrix  $B$ , we refer to  $\ell_T$  as DLAMN (degenerate locally asymptotically mixed normal).

- **Local Asymptotic Brownian Functional structure (LABF)** if

$$(2.15) \quad \ell_{\infty}(\mathbf{u}) = \frac{1}{\sigma} \int_0^1 \mathbf{u}^{\top} G(t) d\mathbf{w}(t) - \frac{1}{2\sigma^2} \int_0^1 \mathbf{u}^{\top} G(t) G^{\top}(t) \mathbf{u} dt,$$

where  $G \in \mathbb{R}^{2 \times 2}$  is an adapted process,  $\mathbf{w} \in \mathbb{R}^2$  is a standard Brownian motion, and the pair  $(G, \mathbf{w})$  is a Gaussian process.

The definitions of LAN, LAMN, and LABF extend to more general likelihood ratios and to any finite number of unknown parameters. For details see [7, Chapter II] (LAN), [3, Chapter 1] (LAMN), and [10, Section 2] (unified approach to LAN, LAMN, and LABF).

The LAN and LAMN properties of  $\ell_T$  imply certain asymptotic efficiency of the corresponding maximum likelihood estimator (MLE):

- (1) If  $\ell_T$  is LAN, then the corresponding MLE is asymptotically efficient in the sense of achieving the lower bound in the Cramer-Rao inequality; for details see [7, Theorem II.12.1].
- (2) If  $\ell_T$  is LAMN, then the corresponding MLE has the maximal concentration property; for details, see [3, Theorem 2.2.1]. Note that the result requires non-degeneracy of the matrix  $B$  and therefore does not immediately extend to DLAMN.

In the LABF case, there are results about asymptotic efficiency of Bayesian estimators [10, Section 3, Proposition 10] and sequential estimators [6, Theorem 2].

To put our results in perspective, let us recall the estimation problem of the drift  $\theta$  in the CAR(1) model, which is the one-dimensional OU process  $Y = Y(t)$  defined by

$$(2.16) \quad dY(t) = \theta Y(t) dt + \sigma dW(t).$$

Here is a summary of the results. For details, see [5, 11].

- The maximum likelihood estimator  $\hat{\theta}_T$  of  $\theta$  using the observations of  $Y(t)$ ,  $0 \leq t \leq T$  is

$$(2.17) \quad \hat{\theta}_T = \frac{\int_0^T Y(t) dY(t)}{\int_0^T Y^2(t) dt};$$

the estimator is strongly consistent as  $T \rightarrow \infty$ :  $\lim_{T \rightarrow \infty} \widehat{\theta}_T = \theta$  with probability one for all  $\theta \in \mathbb{R}$ .

- If  $\theta < 0$  (asymptotically stable or ergodic case), then

$$(2.18) \quad \lim_{T \rightarrow \infty} \sqrt{|\theta|T}(\widehat{\theta}_T - \theta) \stackrel{d}{=} \sqrt{2}|\theta|\xi,$$

where  $\xi$  is a standard normal random variable.

- If  $\theta = 0$  (neutrally stable case), then

$$(2.19) \quad \lim_{T \rightarrow \infty} T(\widehat{\theta}_T - \theta) \stackrel{d}{=} \frac{w^2(1) - 1}{2 \int_0^1 w^2(s) ds},$$

where  $w = w(s)$ ,  $0 \leq s \leq 1$ , is a standard Brownian motion.

- If  $\theta > 0$  (unstable or explosive case), then

$$(2.20) \quad \lim_{T \rightarrow \infty} e^{\theta T}(\widehat{\theta}_T - \theta) \stackrel{d}{=} 2\theta \frac{\eta}{\xi + c},$$

where  $\xi = \sqrt{2\theta} \int_0^\infty e^{-\theta t} dW(t)$  is a standard normal random variable,  $\eta$  is a standard normal random variable independent of  $\xi$ , and  $c = \sqrt{2\theta} Y(0)/\sigma$ . In particular, if  $Y(0) = 0$ , then the limit has the Cauchy distribution and does not depend on  $\sigma$ .

- If  $\theta \neq 0$ , then NLRR holds:

$$(2.21) \quad \lim_{T \rightarrow \infty} \left( \int_0^T Y^2(t) dt \right)^{1/2} (\widehat{\theta}_T - \theta) \stackrel{d}{=} \sigma \eta,$$

where  $\eta$  is a standard normal random variable.

- The normalized log-likelihood ratio

$$\ell_T(u) = \frac{u}{\sigma} \frac{\int_0^T Y(t) dW(t)}{\left( \mathbb{E} \int_0^T Y^2(t) dt \right)^{1/2}} - \frac{u^2}{2\sigma^2} \frac{\int_0^T Y^2(t) dt}{\mathbb{E} \int_0^T Y^2(t) dt}, \quad u \in \mathbb{R}.$$

is LAN if  $\theta < 0$ , LABF if  $\theta = 0$ , and LAMN if  $\theta > 0$ .

Note that (a) the limit distributions in (2.18), (2.19), and (2.21) do not depend on the initial condition; (b) equality (2.21) illustrates the attractive features of NLRR: the rate

$$R(T) = \left( \int_0^T Y^2(t) dt \right)^{1/2}$$

does not explicitly depend on  $\theta$  and the limit distribution does not depend on  $\theta$  or the initial conditions.

Table 1 summarizes the results, where  $F_d(w)$  denotes a generic functional of the standard  $d$ -dimensional Brownian motion and Ch denotes a Cauchy-type distribution (ratio of two independent normal random variables).

Let us now turn to equation (1.1). Asymptotic behavior of estimators (2.4) depends on the roots  $p, q$  of the characteristic equation

$$(2.22) \quad r^2 - \theta_1 r - \theta_2 = 0.$$

TABLE 1. Estimation in CAR(1)

Parameter	Rate	LD	$\ell_T$	NLRR
$\theta < 0$	$\sqrt{T}$	$\mathcal{N}$	LAN	Yes
$\theta = 0$	$T$	$F_1(w)$	LABF	No
$\theta > 0$	$e^{\theta T}$	Ch	LAMN	Yes

There are nine cases to consider:

- (1) The asymptotically stable (ergodic) case, when  $\theta_1 < 0$  and  $\theta_2 < 0$ . All in all, there are three possibilities for the roots:  $q < p < 0$ ,  $q = p < 0$ , or  $p = \lambda + \sqrt{-1}\nu$ ,  $q = \lambda - \sqrt{-1}\nu$ , with  $\lambda < 0$ , but the asymptotic behavior of the estimators is the same in all three cases.
- (2) Six non-ergodic cases with real  $p, q$ :  $q < p = 0$ ,  $q < 0 < p$ ,  $p > q = 0$ ,  $p > q > 0$ ,  $p = q > 0$ , or  $p = q = 0$ .
- (3) The harmonic oscillator, when  $p = \sqrt{-1}\nu$ ,  $q = -\sqrt{-1}\nu$ ,  $\nu > 0$ ;
- (4) Unstable oscillations, when  $p = \lambda + \sqrt{-1}\nu$ ,  $q = \lambda - \sqrt{-1}\nu$ , and  $\lambda > 0, \nu > 0$ .

Table 2 summarizes the results for CAR(2). The detailed statements are in Section 3. In Table 2, we use the same notations as in Table 1. In particular,  $F_d(w)$  denotes a functional of the standard  $d$ -dimensional Brownian motion and Ch is a Cauchy-type distribution.

TABLE 2. Estimation in CAR(2)

Case	$v_1$	$v_2$	LD <sub>1</sub>	LD <sub>2</sub>	NLRR	$\ell_T$
$\theta_1 < 0, \theta_2 < 0$	$\sqrt{T}$	$\sqrt{T}$	$\mathcal{N}$	$\mathcal{N}$	Yes	LAN
$q < 0 < p$	$\sqrt{T}$	$\sqrt{T}$	$\mathcal{N}$	$\mathcal{N}$	Yes	DLAMN
$0 < q < p$	$e^{qT}$	$e^{qT}$	Ch	Ch	Yes	DLAMN
$0 < q = p$	$T^{-1}e^{qT}$	$T^{-1}e^{qT}$	Ch	Ch	Yes	DLAMN
$q < p = 0$	$\sqrt{T}$	$T$	$\mathcal{N}$	$F_1(w)$	Yes ( $\hat{\theta}_{1,T}$ )	LABF/LAN
$q = 0 < p$	$T$	$T$	$F_1(w)$	$F_1(w)$	No	DLAMN
$q = p = 0$	$T$	$T^2$	$F_1(w)$	$F_1(w)$	No	LABF
$\Re(p) = 0$	$T$	$T$	$F_2(w)$	$F_2(w)$	No	LABF
$\Re(p) = \lambda > 0$	$e^{\lambda T}$	$e^{\lambda T}$	Many	Many	Yes	LAMN family

There are obvious similarities between CAR(1) and CAR(2) in the ergodic case and, perhaps less obvious, similarities in the neutrally stable case ( $\theta = 0$  in CAR(1) compared to  $p = \sqrt{-1}\nu$  in CAR(2)) and in the exponentially unstable cases ( $\theta > 0$  in CAR(1) compared to real  $p, q > 0$  in CAR(2)). In both CAR(1) and CAR(2), the rate  $\sqrt{T}$  corresponds to normal distribution in the limit, exponential rate corresponds to a Cauchy-type distribution, and any rate polynomial in  $T$  leads to some functional of the standard Brownian motion. In the case of the positive double root ( $p = q > 0$ ) the rate is slightly slower than exponential, but the limit distribution is still of the Cauchy

type. Also of interest are (a) several appearances of DLAMN instead of LAMN, (b) an unusual combination of asymptotic normality of the estimators and DLAMN (rather than LAN) of  $\ell_T$  when  $q < 0 < p$ , (c) relative compactness rather than convergence in distribution for both the estimators and the normalized log-likelihood ratio when  $p = \lambda + \sqrt{-1}\nu$ ,  $\lambda > 0$ .

It is instructive to compare CAR(2) with  $p = \lambda + \sqrt{-1}\nu$ ,  $\lambda > 0, \nu > 0$ , and the example considered in [16, Section 4.1]:

$$(2.23) \quad \begin{pmatrix} dX_1(t) \\ dX_2(t) \end{pmatrix} = \begin{pmatrix} \lambda & -\nu \\ \nu & \lambda \end{pmatrix} \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} dt + \begin{pmatrix} dW_1(t) \\ dW_2(t) \end{pmatrix}.$$

While the eigenvalues of the matrix in (2.23) are also  $\lambda \pm \sqrt{-1}\nu$ , the special structure of the model ensures that the normalized local log-likelihood ratio is LAMN and the MLEs of  $\lambda$  and  $\nu$ , when normalized by  $\sqrt{2} \lambda e^{\lambda T}$ , converge to a joint limit (which, for zero initial conditions, is the bivariate  $t_2$ -distribution).

### 3. ASYMPTOTIC PROPERTIES OF THE MLE AND THE NORMALIZED LOG-LIKELIHOOD RATIO

Strong consistency of (2.4) is a consequence of a more general result by Basak and Lee [2]. For the sake of completeness, here are the statement and the proof.

**Theorem 3.1.** *With probability one,  $\lim_{T \rightarrow \infty} \hat{\theta}_{1,T} = \theta_1$  and  $\lim_{T \rightarrow \infty} \hat{\theta}_{2,T} = \theta_2$  for every  $\theta \in \mathbb{R}^2$ .*

*Proof.* With  $\hat{\Theta}_T$  defined in (1.5), we find using (1.1) that

$$\hat{\Theta}_T = \begin{pmatrix} 0 & 1 \\ \hat{\theta}_{2,T} & \hat{\theta}_{1,T} \end{pmatrix}.$$

The statement of the theorem now follows from [2, Theorem 2.1 and Remark 3.1].  $\square$

Next, we present the theorem describing the limit distributions of  $\hat{\theta}_{1,T}$  and  $\hat{\theta}_{2,T}$ . The proofs are in Section 5. To keep visual track of the formulas, it is convenient to think of the process  $X$  in (1.1) as a dimensionless quantity and to measure  $t$  in the units  $[t]$  of time. Table 3 summarizes the resulting dimensions of all the variables and parameters in the problem.

TABLE 3. Dimensions in CAR(2)

Quantity	$t$	$X(t)$	$W(t)$	$\dot{X}(t), \theta_1, \hat{\theta}_{1,T}, p, q$	$\sigma$	$\ddot{X}(t), \theta_2, \hat{\theta}_{2,T}$
Units	$[t]$	None	$[t]^{1/2}$	$[t]^{-1}$	$[t]^{-3/2}$	$[t]^{-2}$

One caveat: the auxiliary Brownian motion  $w(s)$ ,  $0 \leq s \leq 1$ , and its parameter  $s$ , as well as all random variables appearing in the limit distributions, are dimensionless.



**Theorem 3.2** (Rate of convergence and limit distributions).

I. ERGODIC CASE. Assume that  $\theta_1 < 0$  and  $\theta_2 < 0$ , and let  $\eta_1, \eta_2$  be iid standard normal random variables. Then

$$(3.1) \quad \lim_{T \rightarrow \infty} \sqrt{T|\theta_1|}(\widehat{\theta}_{1,T} - \theta_1) \stackrel{d}{=} \sqrt{2}|\theta_1|\eta_1, \quad \lim_{T \rightarrow \infty} \sqrt{T|\theta_1|}(\widehat{\theta}_{2,T} - \theta_2) \stackrel{d}{=} \sqrt{2|\theta_2|}|\theta_1|\eta_2.$$

For the rest of the theorem, denote by  $p$  and  $q$  the roots of equation (2.22).

II. NON-ERGODIC CASE: DISTINCT REAL ROOTS. Let  $\xi, \eta$  be iid standard normal random variables, and let  $w = w(s)$ ,  $0 \leq s \leq 1$ , be a standard Brownian motion independent of  $\eta$ .

(A) If  $p > 0$  and  $q < 0$ , then

$$(3.2) \quad \lim_{T \rightarrow \infty} \sqrt{|q|T}(\widehat{\theta}_{1,T} - \theta_1) \stackrel{d}{=} -\frac{1}{p} \lim_{T \rightarrow \infty} \sqrt{|q|T}(\widehat{\theta}_{2,T} - \theta_2) \stackrel{d}{=} \sqrt{2}|q|\eta,$$

(B) If  $p > q > 0$ , then

$$(3.3) \quad \lim_{T \rightarrow \infty} e^{qT}(\widehat{\theta}_{1,T} - \theta_1) \stackrel{d}{=} -\frac{1}{p} \lim_{T \rightarrow \infty} e^{qT}(\widehat{\theta}_{2,T} - \theta_2) \stackrel{d}{=} \frac{2(p+q)q}{p-q} \frac{\eta}{\xi+c},$$

where  $c = \sqrt{2q}(\dot{X}(0) - pX(0))/\sigma$ .

(C) If  $p = 0$  and  $q < 0$ , then  $\theta_1 = q$ ,  $\theta_2 = 0$ , and

$$(3.4) \quad \lim_{T \rightarrow \infty} \sqrt{|\theta_1|T}(\widehat{\theta}_{1,T} - \theta_1) \stackrel{d}{=} \sqrt{2}|\theta_1|\eta, \quad \lim_{T \rightarrow \infty} T\widehat{\theta}_{2,T} \stackrel{d}{=} |\theta_1| \frac{w^2(1) - 1}{2 \int_0^1 w^2(s) ds}.$$

(D) If  $p > 0$  and  $q = 0$ . Then  $\theta_1 = p$ ,  $\theta_2 = 0$ , and

$$(3.5) \quad \lim_{T \rightarrow \infty} \theta_1 T(\widehat{\theta}_{1,T} - \theta_1) \stackrel{d}{=} - \lim_{T \rightarrow \infty} T\widehat{\theta}_{2,T} \stackrel{d}{=} \theta_1 \frac{w^2(1) - 1}{2 \int_0^1 w^2(s) ds}.$$

III. NON-ERGODIC CASE: A DOUBLE ROOT.

(A) If  $p = q > 0$ , then  $\theta_1 = 2q$ ,  $\theta_2 = q^2$ , and

$$(3.6) \quad \lim_{T \rightarrow \infty} \frac{e^{qT}}{qT}(\widehat{\theta}_{1,T} - \theta_1) \stackrel{d}{=} -\frac{1}{q} \lim_{T \rightarrow \infty} \frac{e^{qT}}{qT}(\widehat{\theta}_{2,T} - \theta_2) \stackrel{d}{=} 4\sqrt{2}q \frac{\eta}{\xi+c},$$

where  $c = \sqrt{2q}(\dot{X}(0) - pX(0))/\sigma$  and  $\xi, \eta$  are iid standard normal random variables.

(B) If  $p = q = 0$ , then  $\theta_1 = 0$ ,  $\theta_2 = 0$ , and

$$(3.7) \quad \lim_{T \rightarrow \infty} T\widehat{\theta}_{1,T} \stackrel{d}{=} \frac{2\mathfrak{z}_3(w^2(1) - 1) - 2\mathfrak{z}_1^2(w(1)\mathfrak{z}_1 - \mathfrak{z}_2)}{4\mathfrak{z}_2\mathfrak{z}_3 - \mathfrak{z}_1^4},$$

$$\lim_{T \rightarrow \infty} T^2\widehat{\theta}_{2,T} \stackrel{d}{=} \frac{4\mathfrak{z}_2(w(1)\mathfrak{z}_1 - \mathfrak{z}_2) - \mathfrak{z}_1^2(w^2(1) - 1)}{4\mathfrak{z}_2\mathfrak{z}_3 - \mathfrak{z}_1^4},$$

where  $w = w(s)$ ,  $0 \leq s \leq 1$ , is a standard Brownian motion, and

$$\mathfrak{z}_1 = \int_0^1 w(s) ds, \quad \mathfrak{z}_2 = \int_0^1 w^2(s) ds, \quad \mathfrak{z}_3 = \int_0^1 \left( \int_0^t w(s) ds \right)^2 dt.$$

IV. NON-ERGODIC CASE: COMPLEX ROOTS.

(A) If  $p = \sqrt{-1}\nu$ ,  $\nu > 0$ , then  $\theta_1 = 0$ ,  $\theta_2 = -\nu^2$ , and

$$(3.8) \quad \begin{aligned} \lim_{T \rightarrow \infty} T\widehat{\theta}_{1,T} &\stackrel{d}{=} \frac{2 - w_1^2(1) - w_2^2(1)}{\int_0^1 w_1^2(t)dt + \int_0^1 w_2^2(t)dt}, \\ \lim_{T \rightarrow \infty} T(\widehat{\theta}_{2,T} - \theta_2) &\stackrel{d}{=} 2\nu \frac{\int_0^1 w_1(t)dw_2(t) - \int_0^1 w_2(t)dw_1(t)}{\int_0^1 w_1^2(t)dt + \int_0^1 w_2^2(t)dt}, \end{aligned}$$

where  $w_1, w_2$  are independent standard Brownian motions.

(B) If  $p = \lambda + \sqrt{-1}\nu$ ,  $\lambda > 0$ ,  $\nu > 0$ , then, for  $i = 1, 2$ , the families  $\{e^{\lambda T}(\widehat{\theta}_{i,T} - \theta_i), T > 0\}$  are relatively compact, and all the limit distributions are the form

$$\frac{\bar{\xi}_c \bar{\eta}_c + \bar{\xi}_s \bar{\eta}_s}{\bar{\xi}_c^2 + \bar{\xi}_s^2},$$

where  $(\bar{\xi}_c, \bar{\xi}_s)$  and  $(\bar{\eta}_c, \bar{\eta}_s)$  are independent bivariate normal vectors,  $\mathbb{E}\bar{\eta}_c = \mathbb{E}\bar{\eta}_s = 0$ , and the mean values of  $\bar{\xi}_c$  and  $\bar{\xi}_s$  depend on the initial conditions  $X(0), \dot{X}(0)$ .

One general conclusion of Theorem 3.2 is that, if  $p > 0$  and  $q < p$ , then it is the value of the smaller root  $q$  that determines asymptotic behavior of the estimators. This result comes as a surprise: the asymptotic behavior of both estimators is dictated by the non-dominant mode, even though this mode is “invisible” with probability one. Indeed, the solution of (1.1) is a Gaussian process

$$(3.9) \quad X(t) = X(0)x_1(t) + \dot{X}(0)x_2(t) + \sigma \int_0^t x_2(t-s)dW(s),$$

where the functions  $x_1(t), x_2(t)$  form the fundamental system of solutions for the equation

$$(3.10) \quad \ddot{x}(t) - \theta_1 \dot{x}(t) - \theta_2 x(t) = 0.$$

In other words,  $x_1(0) = 1, \dot{x}_1(0) = 0, x_2(0) = 0, \dot{x}_2(0) = 1$ , and both  $x_1 = x_1(t)$  and  $x_2 = x_2(t)$  satisfy (3.10). The roots of the characteristic equation (2.22) are

$$(3.11) \quad p = \frac{\theta_1 + \sqrt{\theta_1^2 + 4\theta_2}}{2}, \quad q = \frac{\theta_1 - \sqrt{\theta_1^2 + 4\theta_2}}{2}.$$

Then, with the usual modifications for complex,  $p, q$ ,

$$(3.12) \quad x_1(t) = \begin{cases} \frac{qe^{pt} - pe^{qt}}{q - p}, & \text{if } p > q, \\ (1 - qt)e^{qt}, & \text{if } p = q; \end{cases} \quad x_2(t) = \begin{cases} \frac{e^{pt} - e^{qt}}{p - q}, & \text{if } p > q, \\ te^{qt}, & \text{if } p = q. \end{cases}$$

When the roots (3.11) are real and distinct, equation (3.10) has two Lyapunov exponents,  $p$  and  $q$ , and it follows from (3.9) that if  $q < p$ , then

$$\mathbb{P} \left( \lim_{t \rightarrow \infty} \frac{1}{t} \ln |X(t)| = q \right) = \mathbb{P} \left( \lim_{t \rightarrow \infty} \frac{1}{t} \ln |\dot{X}(t)| = q \right) = 0$$

for all initial conditions  $X(0), \dot{X}(0)$ . Thus, Theorem 3.2 shows that if the larger Lyapunov exponent of (3.10) is positive, then the asymptotic behavior of the estimators is determined by the smaller Lyapunov exponent.

As another illustration of the effects of the exponentially unstable mode, note that if  $\theta_2 = 0$ , then (1.1) becomes

$$(3.13) \quad d\dot{X} = \theta_1 \dot{X} + \sigma \dot{W}.$$

In other words,  $\dot{X}$  is a CAR(1) process. If  $\theta_1 < 0$ , then asymptotic behavior of  $\widehat{\theta}_{1,T}$  and  $\widehat{\theta}_{2,T}$  and similar to the CAR(1) situation in ergodic and neutrally stable cases, respectively. If  $\theta_2 \geq 0$ , then there is no clear similarity with CAR(1).

When  $p > 0$ , the normalized limits of  $\widehat{\theta}_{1,T} - \theta_1$  and  $\widehat{\theta}_{2,T} - \theta_2$  are negative multiples of each other. Some correlation is also present when  $p = \lambda + \sqrt{-1}\nu$ ,  $\lambda \geq 0, \nu > 0$ . This type of correlation in non-ergodic multi-parameter models has been observed before; see, for example, [16, Section 4.1]. Still, as the case  $q < 0 = p$  shows, lack of ergodicity does not necessarily imply correlation of the limits.

One can verify qualitative consistency of the results of Theorem 3.2 by considering various limiting regimes for  $p$  and  $q$ . For example, passing to limit  $p \searrow q$  in (3.3) suggests that the rate in the case of the positive double root should be slower than exponential.

Next, we study the possibility of NLRR, that is, existence of a random matrix  $R = R(T)$  such that  $R(T)(\widehat{\theta}_T - \theta)$  converges in distribution to a bivariate normal vector.

**Theorem 3.3** (Normal Limit with a Random Rate (NLRR)). *Denote by  $p$  and  $q$  the roots of equation (2.22). Normal limit with a random rate is possible in the following six cases: ergodic;  $p > q > 0$ ;  $p = q > 0$ ;  $q < 0 < p$ ;  $q < 0 = p$  (for  $\widehat{\theta}_{1,T}$  only);  $p = \lambda + \sqrt{-1}\nu$ ,  $\lambda > 0$ ,  $\nu > 0$ .*

**ERGODIC CASE.** *Assume that  $\theta_1 < 0$ ,  $\theta_2 < 0$ , and let  $\eta_1, \eta_2$  be iid standard normal random variables. Then*

$$(3.14) \quad \begin{aligned} \lim_{T \rightarrow \infty} \left( \int_0^T \dot{X}^2(t) dt \right)^{1/2} (\widehat{\theta}_{1,T} - \theta_1) &\stackrel{d}{=} \sigma \eta_1, \\ \lim_{T \rightarrow \infty} \left( \int_0^T X^2(t) dt \right)^{1/2} (\widehat{\theta}_{2,T} - \theta_2) &\stackrel{d}{=} \sigma \eta_2. \end{aligned}$$

*For the rest of the theorem, denote by  $\eta$  a standard normal random variable.*

**DISTRICT POSITIVE ROOTS.** *Assume that  $p > q > 0$  and define*

$$(3.15) \quad r(T) = \left( \int_0^T (\dot{X}(t) - pX(t))^2 dt \right)^{1/2}.$$

*Then*

$$(3.16) \quad \lim_{T \rightarrow \infty} r(T)(\widehat{\theta}_{1,T} - \theta_1) \stackrel{d}{=} -\frac{1}{p} \lim_{T \rightarrow \infty} r(T)(\widehat{\theta}_{2,T} - \theta_2) \stackrel{d}{=} \frac{p+q}{p-q} \sigma \eta.$$

**POSITIVE DOUBLE ROOT.** *Assume that  $p = q > 0$  and define*

$$r(T) = \frac{1}{T^2} \left( \int_0^T X^2(t) dt \right)^{1/2}.$$

Then

$$(3.17) \quad \lim_{T \rightarrow \infty} r(T)(\widehat{\theta}_{1,T} - \theta_1) \stackrel{d}{=} -\frac{1}{p} \lim_{T \rightarrow \infty} r(T)(\widehat{\theta}_{2,T} - \theta_2) \stackrel{d}{=} 2\sqrt{2}p\sigma\eta.$$

ROOTS OF OPPOSITE SIGN. Assume that  $q < 0 < p$  and let  $r(T)$  be as in (3.15). Then

$$(3.18) \quad \lim_{T \rightarrow \infty} r(T)(\widehat{\theta}_{1,T} - \theta_1) \stackrel{d}{=} -\frac{1}{p} \lim_{T \rightarrow \infty} r(T)(\widehat{\theta}_{2,T} - \theta_2) \stackrel{d}{=} \sigma\eta.$$

ZERO ROOT. Assume that  $q < 0$ ,  $p = 0$ , and define

$$r(T) = T^{-3/2} \int_0^T X^2(t) dt.$$

Then

$$(3.19) \quad \lim_{T \rightarrow \infty} r(T)(\widehat{\theta}_{1,T} - \theta_1) \stackrel{d}{=} \frac{\sigma^2}{\sqrt{2}|q|^{3/2}} \eta.$$

COMPLEX ROOTS. Assume that  $p = \lambda + \sqrt{-1}\nu$ ,  $\lambda > 0$ ,  $\nu > 0$  and define matrices

$$A_T = \begin{pmatrix} \nu & 0 \\ \lambda & -1 \end{pmatrix} e^{-\lambda T}, \quad B(x, y) = \frac{1}{x^2 + y^2} \begin{pmatrix} x & y \\ -y & x \end{pmatrix}.$$

Then there exists a bivariate normal vector  $(\mathbf{u}_s, \mathbf{u}_c)$  such that the family

$$B(\mathbf{u}_s, \mathbf{u}_c) A_T \Psi_T(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}), \quad T > 0,$$

is relatively compact, and all partial limits are bivariate normal random vectors independent of  $(\mathbf{u}_s, \mathbf{u}_c)$ .

Theorem 3.3 suggests two negative conclusions: (a)  $\Psi_T^{1/2}$  is usually not the correct random normalization (which is especially striking when  $p = \lambda + \sqrt{-1}\nu$ ); (b) existence of NLRR in non-ergodic CAR(2) is not as helpful as in CAR(1), because the rates and/or the limit distributions depend on the unknown parameters.

Finally, we describe the asymptotic structure of the normalized log-likelihood ratio (2.11).

**Theorem 3.4** (The structure of  $\ell_T$ ). Denote by  $p$  and  $q$  the roots of equation (2.22), by  $\text{diag}(x, y)$  the diagonal matrix with  $x$  and  $y$  on the main diagonal, and by  $\mathbf{b}_p$  the column vector

$$\mathbf{b}_p = \begin{pmatrix} 1 \\ p \end{pmatrix}.$$

If  $\theta_1 < 0$ ,  $\theta_2 < 0$  (ergodic case), and  $A_T = \text{diag}(T^{-1/2}, T^{-1/2})$ , then  $\ell_T$  is LAN.

If  $p > 0$ ,  $q < p$ , and  $A_T = e^{-pT} \mathbf{b}_p \mathbf{b}_p^\top$ , then  $\ell_T$  is degenerate LAMN and the matrix  $B$  in (2.14) is

$$(3.20) \quad B = \frac{(1+p^2)^2}{2p} \sigma^2 \zeta^2 \mathbf{b}_p \mathbf{b}_p^\top,$$

where  $\zeta$  is a standard normal random variable.

If  $q < 0 = p$  and  $A_T = \text{diag}(T^{-1}, T^{-1/2})$  then  $\ell_T$  is mixed LABF/LAN:

$$\ell_\infty(\mathbf{u}) = \mathbf{u}^\top \boldsymbol{\xi} - \frac{1}{2} \mathbf{u}^\top B \mathbf{u},$$

where

$$\boldsymbol{\xi} = \begin{pmatrix} \sigma |q|^{-1} \int_0^1 w(s) dw(s) \\ (2|q|)^{-1/2} \sigma \eta \end{pmatrix}, \quad B = \begin{pmatrix} \sigma^2 |q|^{-2} \int_0^1 w^2(s) ds & 0 \\ 0 & \sigma^2 (2|q|)^{-1} \end{pmatrix},$$

$\eta$  is a standard normal random variable,  $w$  is a standard Brownian motion, and  $\eta$  and  $w$  are independent.

If  $p = q > 0$  and  $A_T = T^{-1} e^{-pT} \mathbf{b}_p \mathbf{b}_p^\top$ , then  $\ell_T$  is degenerate LAMN and the matrix  $B$  in (2.14) is given by (3.20).

If  $p = q = 0$  and  $A_T = \text{diag}(T^{-2}, T^{-1})$ , then  $\ell_T$  is LABF and the matrix  $G(t)$  in (2.15) is

$$G(t) = \begin{pmatrix} \sigma \int_0^t w(s) ds & 0 \\ \sigma w(t) & 0 \end{pmatrix},$$

where  $w$  is a standard Brownian motions.

If  $p = \sqrt{-1}\nu$ ,  $\nu > 0$ , and  $A_T = \text{diag}(T^{-1}, T^{-1})$ , then  $\ell_T$  is LABF and the matrix  $G(t)$  in (2.15) is

$$G(t) = \begin{pmatrix} \sigma w_1(t) & \sigma w_2(t) \\ -\sigma w_2(t) & \sigma w_1(t) \end{pmatrix},$$

where  $w_1$  and  $w_2$  are independent standard Brownian motions.

If  $p = \lambda + \sqrt{-1}\nu$ ,  $\lambda, \nu > 0$ , and

$$A_T = \begin{pmatrix} \nu & 0 \\ \lambda & -1 \end{pmatrix} e^{-\lambda T},$$

then  $\{\ell_T, T > 0\}$  is a relatively compact LAMN family having all partial limits of the form (2.14).

One general conclusion of Theorem 3.4 is that, if  $p > 0$  and  $q < p$ , then it is the value of the larger root  $p$  that determines asymptotic behavior of the normalized log-likelihood ratio, which is in sharp contrast with Theorem 3.2.

## 4. PREPARATION FOR THE PROOFS

To study asymptotic behavior of  $\widehat{\theta}_{1,T}$  and  $\widehat{\theta}_{2,T}$ , we need the expressions for the residuals  $\widehat{\theta}_{i,T} - \theta_i$ ,  $i = 1, 2$ :

$$(4.1) \quad \begin{aligned} \widehat{\theta}_{1,T} - \theta_1 &= \frac{\left(\int_0^T X^2(t)dt\right) \left(\int_0^T \dot{X}(t)\sigma dW(t)\right) - \left(\int_0^T X(t)\dot{X}(t)dt\right) \left(\int_0^T X(t)\sigma dW(t)\right)}{\left(\int_0^T \dot{X}^2(t)dt\right) \left(\int_0^T X^2(t)dt\right) - \left(\int_0^T X(t)\dot{X}(t)dt\right)^2}, \\ \widehat{\theta}_{2,T} - \theta_2 &= \frac{\left(\int_0^T \dot{X}^2(t)dt\right) \left(\int_0^T X(t)\sigma dW(t)\right) - \left(\int_0^T X(t)\dot{X}(t)dt\right) \left(\int_0^T \dot{X}(t)\sigma dW(t)\right)}{\left(\int_0^T \dot{X}^2(t)dt\right) \left(\int_0^T X^2(t)dt\right) - \left(\int_0^T X(t)\dot{X}(t)dt\right)^2}. \end{aligned}$$

These equalities follow directly from (2.4) and (1.1).

Equation (1.1) has a closed-form solution (3.9), meaning that (4.1) can be written in terms of integrals of the type  $\int_0^t f(s)dW(s)$ . Unfortunately, this direct approach quickly leads to intractable expressions. Therefore, despite availability of explicit formulas, a more sophisticated approach to the analysis of (4.1) is necessary. In the ergodic case, the ergodic theorem provides all the necessary tools, and when  $\theta_1 = \theta_2 = 0$ , the expressions are simplified using self-similarity of the standard Brownian motion. Other cases benefit from the following construction.

Given two square-integrable on  $[0, T]$  functions  $f, g$ , define

$$(4.2) \quad \begin{aligned} N(T; f, g) &= \left(\int_0^T f^2(t)dt\right) \left(\int_0^T g(t)\sigma dW(t)\right) \\ &\quad - \left(\int_0^T f(t)g(t)dt\right) \left(\int_0^T f(t)\sigma dW(t)\right), \end{aligned}$$

$$(4.3) \quad D(T; f, g) = \left(\int_0^T f^2(t)dt\right) \left(\int_0^T g^2(t)dt\right) - \left(\int_0^T f(t)g(t)dt\right)^2.$$

Clearly,  $D(T; f, g) = D(T; g, f)$ , but in general  $N(T; f, g) \neq N(T; g, f)$ . Then formulas (4.1) become

$$(4.4) \quad \widehat{\theta}_{1,T} - \theta_1 = \frac{N(T; X, \dot{X})}{D(T; X, \dot{X})}, \quad \widehat{\theta}_{2,T} - \theta_2 = \frac{N(T; \dot{X}, X)}{D(T; X, \dot{X})}.$$

The general idea of the proof of Theorem 3.2 is to find the asymptotic behavior of  $D(T; X, \dot{X})$ ,  $N(T; X, \dot{X})$ , and  $N(T; \dot{X}, X)$ , as  $T \rightarrow \infty$ . To keep track of the results, note that, according to Table 3,  $D(T; X, \dot{X})$  is dimensionless,  $N(T; X, \dot{X})$  is measured in  $[t]^{-1}$ , and  $N(T; \dot{X}, X)$  is measured in  $[t]^{-2}$ .

For every real numbers  $\alpha, \beta, \gamma, \kappa$  and every square-integrable functions  $f, g$ , we have the following identities:

$$(4.5) \quad D(T; \alpha f + \beta g, \gamma f + \kappa g) = (\alpha\kappa - \beta\gamma)^2 D(T; f, g),$$

$$(4.6) \quad N(T; \alpha f + \beta g, \gamma f + \kappa g) = (\alpha^2\kappa - \alpha\beta\gamma)N(T; f, g) + (\beta^2\gamma - \alpha\beta\kappa)N(T; g, f).$$

Recall that  $0_{\text{a.s.}}(t)$  denotes a continuous random process converging to zero with probability one as  $t \rightarrow \infty$ . We will often use the following result: if  $f(t) > 0$  is a

continuous process,  $F(T) = \int_0^T f(s)ds$ , and  $\lim_{T \rightarrow \infty} F(T) = \infty$  with probability one, then

$$(4.7) \quad \frac{1}{F(T)} \int_0^T f(t) 0_{\text{a.s.}}(t) dt = 0_{\text{a.s.}}(T).$$

Indeed, (4.7) is immediate if the integral  $\int_0^\infty f(t) 0_{\text{a.s.}}(t) dt$  converges; otherwise, (4.7) follows after one application of L'Hospital's rule.

Here are some other technical results to be used later. For  $r > 0$ , define Gaussian random variables

$$(4.8) \quad \xi_r = \int_0^\infty e^{-rs} dW(s), \quad \eta_r(T) = \int_0^T e^{-r(T-s)} dW(s).$$

We have  $\mathbb{E}\eta_r(T) = 0$ ,  $\mathbb{E}\eta_r^2(T) = (1 - e^{-2rT})/(2r)$ , and therefore

$$(4.9) \quad \lim_{T \rightarrow \infty} \eta_r(T) \stackrel{d}{=} \eta_r,$$

where the random variable  $\eta_r$  is normal with mean zero and variance  $1/(2r)$ . Since

$$(4.10) \quad \mathbb{E}\xi_q \eta_r(T) = e^{-rT} \int_0^T e^{(r-q)s} ds \rightarrow 0, \quad T \rightarrow \infty,$$

it follows that  $\eta_r$  and  $\xi_q$  are independent for every  $q, r > 0$ .

A more sophisticated version of the above observations is the following result.

**Proposition 4.1.** *Let  $M = (M_1(t), \dots, M_d(t))$ ,  $0 \leq t \leq 1$ , be a  $d$ -dimensional continuous Gaussian martingale with  $M(0) = 0$ , and let  $M_T = (M_{T,1}(t), \dots, M_{T,d}(t))$ ,  $T \geq 0$ ,  $0 \leq t \leq 1$ , be a family of continuous square-integrable  $d$ -dimensional martingales such that  $M_T(0) = 0$  for all  $T$  and, for every  $t \in [0, 1]$  and  $i, j = 1, \dots, d$ ,*

$$\lim_{T \rightarrow \infty} \langle M_{T,i}, M_{T,j} \rangle(t) = \langle M_i, M_j \rangle(t)$$

*in probability. Then  $\lim_{T \rightarrow \infty} M_T \stackrel{\mathcal{L}}{=} M$  in the topology of continuous functions on  $[0, 1]$ .*

*Proof.* Modulo a non-essential (in this case) difference between a sequence and a family indexed by the positive reals, this is a particular case of Theorem VIII.3.11 in [8].  $\square$

For every  $r > 0$ , the process  $\eta_r(t)$ ,  $t \geq 0$ , is ergodic (in fact, strictly mixing). Therefore, the ergodic theorem implies

$$(4.11) \quad \frac{1}{T} \int_0^T \eta_r(t) dt = 0_{\text{a.s.}}(T), \quad \frac{1}{T} \int_0^T \eta_r^2(t) dt = \frac{1}{2r} + 0_{\text{a.s.}}(T),$$

and, together with Proposition 4.1

$$(4.12) \quad \begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{\sqrt{T}} \int_0^T \eta_r(t) dW(t) &\stackrel{d}{=} \eta_r^\perp, \\ \lim_{T \rightarrow \infty} \left( \int_0^T \eta_r^2(t) dt \right)^{-1/2} \int_0^T \eta_r(t) dW(t) &\stackrel{d}{=} \sqrt{2r} \eta_r^\perp, \end{aligned}$$

where  $\eta_r^\perp$  is normal with mean zero and variance  $1/(2r)$ , and the random variables  $(\xi_q, \eta_r, \eta_r^\perp)$  are jointly independent for every  $q, r > 0$ .

Asymptotic analysis of certain stochastic integrals can benefit from the law of iterated logarithm. Recall that if  $f$  is locally square-integrable adapted process,  $M(t) = \int_0^t f(s)dW(s)$  and  $\langle M \rangle(t) = \int_0^t f^2(s)ds \nearrow +\infty$ ,  $t \rightarrow \infty$ , with probability one, then there exists a standard Brownian motion  $\bar{W}$  such that

$$M(t) = \bar{W}(\langle M \rangle(t)).$$

The law of iterated logarithm for  $\bar{W}$  implies

$$(4.13) \quad \lim_{T \rightarrow \infty} \frac{M(T)}{\sqrt{\langle M \rangle(T)} (\ln \langle M \rangle(T))^\varepsilon} = 0_{\text{a.s.}}(T)$$

for every  $\varepsilon > 0$ . For example, if  $\varepsilon > 0$ , then

$$(4.14) \quad T^{-\varepsilon} \eta_r(T) = 0_{\text{a.s.}}(T).$$

To conclude the general discussion we establish an integration by parts formula. As a motivation, recall that analysis of CAR(1) in the exponentially unstable case leads to the function  $V_p(t) = \int_0^t e^{p(t-s)} dW(s)$ ,  $p > 0$ , for which integration by parts shows that

$$(4.15) \quad e^{-pT} \left( \int_0^T V_p(t) dW(t) - e^{pT} \xi_p \eta_p(T) \right) = 0_{\text{a.s.}}(T).$$

Since exponentially unstable solutions of equation (3.10) are of the form  $e^{pt} f(t)$ , where the function  $f$  grows at most polynomially, we generalize (4.15) as follows.

**Proposition 4.2.** *Given a deterministic (for simplicity) and locally square-integrable function  $f$ , define  $S_f(t) = \int_0^t f(s)dW(s)$ .*

*Let functions  $\varphi$  and  $\psi$  be such that*

$$(4.16) \quad \int_0^{+\infty} \psi^2(t) dt = +\infty, \quad e^{pT} |\varphi(T)| + e^{-pT} |\psi(T)| \leq C(1 + T^r), \quad T \geq 0,$$

*for some  $p, C, r > 0$ . Then*

$$(4.17) \quad e^{-qT} \left( \int_0^T S_\varphi(t) dS_\psi(t) - S_\varphi(T) S_\psi(T) \right) = 0_{\text{a.s.}}(T) \quad \text{for all } q > 0.$$

*Proof.* By the Itô formula,

$$(4.18) \quad S_\varphi(T) S_\psi(T) - \int_0^T S_\varphi(t) dS_\psi(t) = \int_0^T \varphi(t) \psi(t) dt + \int_0^T S_\psi(t) dS_\varphi(t),$$

and it follows from (4.16) that

$$e^{-qT} \int_0^T |\varphi(t) \psi(t)| dt \leq C_1 e^{-qT} (1 + T^{r+1}) \rightarrow 0, \quad T \rightarrow \infty.$$

To estimate the second term on the right-hand side of (4.18), recall that, if  $M = M(t)$  is a continuous square-integrable martingale, then, by the strong law of large numbers, a finite limit

$$\lim_{T \rightarrow \infty} \frac{M(T)}{1 + \langle M \rangle(T)}$$

exists with probability one ([14, Corollary 2 to Theorem 2.6.10]). As a result, if  $F = F(t)$  is a function such that  $F(T) \langle M \rangle(T) = 0_{\text{a.s.}}(T)$ , then  $F(T) M(T) = 0_{\text{a.s.}}(T)$ .



Next, consider

$$N(t) = \int_0^t S_\psi(s) dS_\varphi(s) = \int_0^t S_\psi(s) \varphi(s) dW(s) \quad \text{with} \quad \langle N \rangle(t) = \int_0^t S_\psi^2(s) \varphi^2(s) ds.$$

We need to show that  $e^{-qT} \langle N \rangle(T) = 0_{\text{a.s.}}(T)$ . By (4.16),

$$(4.19) \quad \langle S_\psi \rangle(t) = \int_0^t \psi^2(s) ds \leq C_2 e^{2pt} (1 + t^{2r}),$$

and therefore

$$(4.20) \quad \lim_{t \rightarrow \infty} e^{-qt} \langle S_\psi \rangle(t) \varphi^2(t) \leq \lim_{t \rightarrow \infty} C_3 e^{-qt} (1 + t^{3r}) = 0, \quad q > 0.$$

By assumption,  $\sup_t \langle S_\psi \rangle(t) = \infty$ . We then use (4.13) and (4.19) to conclude that

$$\lim_{T \rightarrow \infty} \frac{Q^2(T)}{\langle S_\psi \rangle(T) T^\varepsilon} = 0, \quad \varepsilon > 0.$$

Therefore,

$$\lim_{T \rightarrow \infty} e^{-qT} S_\psi^2(T) \varphi^2(T) = \lim_{T \rightarrow \infty} T^\varepsilon e^{-qT} \langle S_\psi \rangle(T) \varphi^2(T) \frac{Q^2(T)}{\langle S_\psi \rangle(T) T^\varepsilon} = 0,$$

which, by L'Hospital's rule, implies  $\lim_{T \rightarrow \infty} e^{-pT} \langle N \rangle(T) = 0_{\text{a.s.}}(T)$ .

*This completes the proof of Proposition 4.2.* □

We will use Proposition 4.2 with  $q = p$  to simplify various stochastic integrals. As a quick illustration, let us verify (4.15). Take  $\varphi(t) = e^{-pt}$ ,  $\psi(t) = e^{pt}$ . Then, together with (4.8), equality (4.17) implies

$$\begin{aligned} e^{-pT} \int_0^T \left( \int_0^t e^{p(t-s)} dW(s) \right) dW(t) &= \left( \int_0^T e^{-p(T-t)} dW(t) \right) \left( \int_0^T e^{-pt} dW(t) \right) \\ &+ 0_{\text{a.s.}}(T) = \eta_p(T) \xi_p + 0_{\text{a.s.}}(T). \end{aligned}$$

## 5. PROOFS OF THEOREMS 3.2–3.4

### Proof of Theorem 3.2.

I. THE ERGODIC CASE. See [1, Remark 2 after Theorem 4.6.2].

II. NON-ERGODIC CASE: DISTINCT REAL ROOTS. If  $p \neq q$ , then (3.9) and (3.12) imply

$$(5.1) \quad X(t) = V_p(t) - V_q(t), \quad \dot{X}(t) = pV_p(t) - qV_q(t),$$

where

$$(5.2) \quad \begin{aligned} V_p(t) &= e^{pt} U_p(t), \quad U_p = \frac{\dot{X}(0) - X(0)q}{p - q} + \frac{\sigma}{p - q} \int_0^t e^{-ps} dW(s), \\ V_q(t) &= e^{qt} U_q(t), \quad U_q = \frac{\dot{X}(0) - X(0)p}{p - q} + \frac{\sigma}{p - q} \int_0^t e^{-qs} dW(s). \end{aligned}$$

By (4.5) and (4.6)

$$(5.3) \quad \begin{aligned} D(T; X, \dot{X}) &= (p - q)^2 D(T; V_p, V_q), \\ N(T; X, \dot{X}) &= (p - q) \left( N(T; V_p, V_q) + N(T; V_q, V_p) \right), \\ N(T; \dot{X}, X) &= -(p - q) \left( p N(T; V_p, V_q) + q N(T; V_q, V_p) \right). \end{aligned}$$

To complete the proof, it now remains to use the specific expressions for the functions  $V_p$  and  $V_q$ , which we do next.

II(A). **ROOTS OF OPPOSITE SIGN:**  $q < 0 < p$ . Using notations (4.8), define Gaussian random variable  $\zeta_p$  by

$$\zeta_p = \frac{\dot{X}(0) - X(0)q}{p - q} + \frac{\sigma}{p - q} \int_0^{+\infty} e^{-ps} dW(s) = \frac{\dot{X}(0) - X(0)q}{p - q} + \frac{\sigma}{p - q} \xi_p$$

and note that

$$V_p(t) = e^{pt} (\zeta_p + 0_{\text{a.s.}}(t)), \quad V_q(t) = \frac{\sigma}{p - q} \eta_{|q|}(t) + \frac{\dot{X}(0) - X(0)p}{p - q} e^{qt}.$$

Then

$$\begin{aligned} \int_0^T V_p^2(t) dt &= \left( \frac{\zeta_p^2}{2p} + 0_{\text{a.s.}}(T) \right) e^{2pT}, \quad \int_0^T V_q^2(t) dt = \frac{\sigma^2 T}{2|q|(p - q)^2} (1 + 0_{\text{a.s.}}(T)), \\ \int_0^T V_p(t) V_q(t) dt &= \int_0^T e^{pt} (\zeta_p + 0_{\text{a.s.}}(t)) V_q(t) dt = e^{pT} (\zeta_p \eta_{|q|}(T) + 0_{\text{a.s.}}(T)), \\ \int_0^T V_p(t) dW(t) &= \sqrt{T} e^{pT} 0_{\text{a.s.}}(T), \quad \lim_{T \rightarrow \infty} \frac{1}{\sqrt{T}} \int_0^T V_q(t) dW(t) \stackrel{d}{=} \frac{\sigma}{p - q} \eta_{|q|}^\perp. \end{aligned}$$

Plugging the results into (5.3),

$$\begin{aligned} D_T(T; X, \dot{X}) &= (T e^{2pT}) \left( \frac{\zeta_p^2}{2p} \frac{\sigma^2}{2|q|} + 0_{\text{a.s.}}(T) \right), \\ \lim_{T \rightarrow \infty} \frac{e^{-2pT}}{\sqrt{T}} N(T; V_p, V_q) &\stackrel{d}{=} \frac{\sigma^2}{p - q} \frac{\zeta_p^2}{2p} \eta_{|q|}^\perp, \quad \lim_{T \rightarrow \infty} \frac{e^{-2pT}}{\sqrt{T}} N(T; V_q, V_p) \stackrel{d}{=} 0. \end{aligned}$$

Then both equalities in (3.2) follow from (4.4).

II(B). **DISTINCT POSITIVE ROOTS:**  $0 < q < p$ . When  $p > q > 0$ , computations are very similar to the case  $q < 0 < p$ . The difference comes from the fact that, for  $q > 0$ , we have

$$V_q(t) = e^{qt} \left( \frac{\sigma}{p - q} \xi_q + \frac{\dot{X}(0) - X(0)p}{p - q} + 0_{\text{a.s.}}(t) \right).$$

Using Proposition 4.2 and notations

$$\zeta_p = \frac{\dot{X}(0) - X(0)q}{p - q} + \frac{\sigma}{p - q} \xi_p, \quad \zeta_q = \frac{\dot{X}(0) - X(0)p}{p - q} + \frac{\sigma}{p - q} \xi_q,$$

we find

$$(5.4) \quad D(T; X, \dot{X}) = \frac{\zeta_p^2 \zeta_q^2 (p-q)^4}{4pq(p+q)^2} e^{2(p+q)T},$$

$$\lim_{T \rightarrow \infty} e^{-(q+2p)T} N(T; V_p, V_q) \stackrel{d}{=} \sigma \zeta_p^2 \zeta_q \left( \frac{\eta_q}{2p} - \frac{\eta_p}{p+q} \right),$$

$$\lim_{T \rightarrow \infty} e^{-(q+2p)T} N(T; V_q, V_p) \stackrel{d}{=} 0.$$

It remains to observe that

$$\frac{\eta_q}{2p} - \frac{\eta_p}{p+q}$$

is a Gaussian random variable, independent of  $(\zeta_p, \zeta_q)$ , with mean zero and variance  $(p-q)^2/(8p^2q(p+q)^2)$ . Then both equalities in (3.3) follow from (4.4).

II(C). LARGER ROOT IS ZERO:  $q < p = 0$ . By (5.1) with  $p = 0$ ,

$$(5.5) \quad X(t) = V_0(t) - V_q(t), \quad \dot{X}(t) = -qV_q(t),$$

and

$$V_0(t) = U_0(t) = \frac{\dot{X}(0) - X(0)q}{|q|} + \frac{\sigma}{|q|} W(t), \quad V_q = \frac{\dot{X}(0)}{|q|} e^{qt} + \frac{\sigma}{|q|} \eta_{|q|}(t).$$

Using (4.11), (4.12), and (4.14), we find:

$$(5.6) \quad \int_0^T V_q^2(t) dt = \frac{\sigma^2 T}{2|q|^3} (1 + 0_{\text{a.s.}}(T)),$$

$$(5.7) \quad \int_0^T V_0^2(t) dt = \frac{\sigma^2 T^2}{q^2} \left( \frac{1}{T^2} \int_0^T W^2(t) dt + 0_{\text{a.s.}}(T) \right),$$

$$(5.8) \quad \int_0^T V_0(t) \dot{X}(t) dt = T^{3/2} 0_{\text{a.s.}}(T),$$

$$(5.9) \quad \lim_{T \rightarrow \infty} \frac{1}{\sqrt{T}} \int_0^T V_q(t) dW(t) \stackrel{d}{=} \frac{\sigma}{|q|} \eta_{|q|}^\perp,$$

$$(5.10) \quad \int_0^T U_0(t) dW(t) = \frac{\sigma T}{|q|} \left( \frac{W^2(T) - T}{2T} + 0_{\text{a.s.}}(T) \right).$$

Self-similarity of the standard Brownian motion implies

$$(5.11) \quad \frac{W^2(T) - T}{2T} \stackrel{d}{=} \frac{w^2(1) - 1}{2}, \quad \frac{1}{T^2} \int_0^T W^2(t) dt \stackrel{d}{=} \int_0^1 w^2(s) ds.$$

Combining (5.6)–(5.11) with (5.3) yields

$$D(T; X, \dot{X}) = \frac{\sigma^4 T^3}{2|q|^3} \left( \frac{1}{T^2} \int_0^T W^2(t) dt + 0_{\text{a.s.}}(T) \right),$$

$$\lim_{T \rightarrow \infty} T^{-5/2} N(T; X, \dot{X}) \stackrel{d}{=} \frac{\sigma^4}{q^2} \eta_{|q|}^\perp \int_0^1 w^2(s) ds,$$

$$\lim_{T \rightarrow \infty} T^{-2} N(T; \dot{X}, X) \stackrel{d}{=} \frac{\sigma^4}{4q^2} (w^2(1) - 1).$$

Then both equalities in (3.4) follow from (4.4).

To show independence of  $\eta$  and  $w$ , take a standard Brownian motion  $\tilde{w} = \tilde{w}(t)$ ,  $t \in [0, 1]$ , that is independent of  $w$  and apply Proposition 4.1 with

$$M(t) = (w(t), \tilde{w}(t)), \quad M_T(t) = \left( \frac{1}{\sqrt{T}} \int_0^{Tt} dW(s), \frac{\sqrt{2|q|}}{\sqrt{T}} \int_0^{Tt} \eta_{|q|}(s) dW(s) \right).$$

II(D). SMALLER ROOT IS ZERO:  $q = 0 < p$ . We have

$$\begin{aligned} X(t) &= V_p(t) - V_0(t), \quad \dot{X}(t) = pV_p(t), \\ V_0(t) &= U_0(t) = \frac{\dot{X}(0) - X(0)p}{p} + \frac{\sigma}{p}W(t), \quad V_p(t) = e^{pt}(\zeta_p + 0_{\text{a.s.}}(t)), \\ \zeta_p &= \frac{\dot{X}(0)}{p} + \frac{\sigma}{p} \int_0^\infty e^{-pt} dW(t). \end{aligned}$$

Then

$$\begin{aligned} \int_0^T V_p^2(t) dt &= \frac{e^{2pT}}{2p} (\zeta_p^2 + 0_{\text{a.s.}}(T)), \quad \int_0^T V_0^2(t) dt = \frac{\sigma^2 T^2}{p^2} \left( \frac{1}{T^2} \int_0^T W^2(t) dt + 0_{\text{a.s.}}(T) \right), \\ \int_0^T V_0(t) V_p(t) dt &= T e^{pT} 0_{\text{a.s.}}(T), \quad \int_0^T V_p(t) dW(t) = e^{pT} (\zeta_p \eta_p(T) + 0_{\text{a.s.}}(T)), \\ \int_0^T U_0(t) dW(t) &= \frac{\sigma T}{p} \left( \frac{W^2(T) - T}{2T} + 0_{\text{a.s.}}(T) \right). \end{aligned}$$

By (5.3),

$$\begin{aligned} D(T; X, \dot{X}) &= \frac{\sigma^2 T^2 e^{2pT}}{2p} \left( \zeta_p^2 \frac{1}{T^2} \int_0^T W^2(t) dt + 0_{\text{a.s.}}(T) \right), \\ N(T; X, \dot{X}) &= \frac{\sigma^2}{2p} T e^{2pT} \left( \zeta_p^2 \frac{W^2(T) - T}{2T} + 0_{\text{a.s.}}(T) \right), \\ N(T; \dot{X}, X) &= -\frac{\sigma^2}{2} T e^{2pT} \left( \zeta_p^2 \frac{W^2(T) - T}{2T} + 0_{\text{a.s.}}(T) \right), \end{aligned}$$

and then both equalities in (3.5) follow from (4.4).

III(A). POSITIVE DOUBLE ROOT:  $p = q > 0$ . With  $x_1(t) = (1 - qt)e^{qt}$ ,  $x_2(t) = te^{qt}$ , (3.9) becomes

$$X(t) = X(0)(1 - qt)e^{qt} + \dot{X}(0)te^{qt} + \sigma \int_0^t (t - s)e^{q(t-s)} dW(s), \quad \dot{X}(t) = qX(t) + Q(t),$$

where

$$Q(t) = \left( \dot{X}(0) - X(0)q + \sigma \int_0^t e^{-qs} dW(s) \right) e^{qt}.$$

If we define

$$\zeta = \dot{X}(0) - X(0)q + \sigma \int_0^{+\infty} e^{-qs} dW(s),$$

then

$$X(t) = te^{qt}(\zeta + 0_{\text{a.s.}}(t)), \quad Q(t) = e^{qt}(\zeta + 0_{\text{a.s.}}(t)).$$

By (4.5),

$$(5.12) \quad D(T; X, \dot{X}) = e^{4qT} \left( \frac{\zeta^4}{16q^4} + 0_{\text{a.s.}}(T) \right);$$

note that the same result follows after passing to the limit  $p \searrow q$  in (5.4).

Next, define

$$\eta_{q,1}(t) = \int_0^t e^{-q(t-s)} dW(s), \quad \eta_{q,2}(t) = \frac{1}{t} \int_0^t s e^{-q(t-s)} dW(s),$$

and observe that

$$(5.13) \quad \lim_{T \rightarrow \infty} T(\eta_{q,1}(T) - \eta_{q,2}(T)) \stackrel{d}{=} \frac{\eta}{2q^{3/2}},$$

where  $\eta$  is a standard Gaussian random variable, independent of  $\zeta$ .

Therefore,

$$\begin{aligned} \int_0^T X^2(t) dt &= T^2 e^{2qT} \left( \frac{\zeta^2}{2q} + 0_{\text{a.s.}}(T) \right), & \int_0^T Q^2(t) dt &= e^{2qT} \left( \frac{\zeta^2}{2q} + 0_{\text{a.s.}}(T) \right), \\ \int_0^T X(t)Q(t) dt &= T e^{2qt} \left( \frac{\zeta^2}{2q} + 0_{\text{a.s.}}(T) \right), & \int_0^T Q(t) dW(t) &= e^{qT} \left( \zeta \eta_{q,1}(T) + 0_{\text{a.s.}}(T) \right), \\ & & \int_0^T X(t) dW(t) &= T e^{qT} \left( \zeta \eta_{q,2}(T) + 0_{\text{a.s.}}(T) \right). \end{aligned}$$

To continue,

$$\begin{aligned} N(T; X, Q) &= \frac{\sigma}{2q} T^2 e^{3qT} (\eta_{q,1}(T) - \eta_{q,2}(T)) (\zeta^3 + 0_{\text{a.s.}}(T)), \\ N(T; Q, X) &= \frac{\sigma}{2q} T e^{3qT} (\eta_{q,2}(T) - \eta_{q,1}(T)) (\zeta^3 + 0_{\text{a.s.}}(T)). \end{aligned}$$

It remains to observe that

$$N(T; X, \dot{X}) = N(T; X, Q), \quad N(T; \dot{X}, X) = -qN(T; X, Q) + N(T; Q, X).$$

Then both equalities in (3.6) follow from (4.4) and (5.13).

III(B). ZERO DOUBLE ROOT:  $p = q = 0$ . In this case the result follows directly from (4.1) using self-similarity of the standard Brownian motion:  $W(T \cdot) \stackrel{\mathcal{L}}{=} \sqrt{T} w(\cdot)$ . Recall the notations

$$\mathfrak{z}_1 = \int_0^1 w(s) ds, \quad \mathfrak{z}_2 = \int_0^1 w^2(s) ds, \quad \mathfrak{z}_3 = \int_0^1 \left( \int_0^t w(s) ds \right)^2 dt.$$

With  $\theta_1 = \theta_2 = 0$ , equation (1.1) becomes  $d\dot{X} = \sigma dW$ , and therefore

$$\dot{X}(t) = \dot{X}(0) + \sigma W(t), \quad X(t) = X(0) + \dot{X}(0)t + \sigma \int_0^t W(s) ds.$$

Then

$$\begin{aligned} D(T; X, \dot{X}) &\stackrel{d}{=} \frac{\sigma^4 T^6}{4} \left( 4\mathfrak{z}_2\mathfrak{z}_3 - \mathfrak{z}_1^4 + 0_{\text{a.s.}}(T) \right), \\ N(T; X, \dot{X}) &\stackrel{d}{=} \frac{\sigma^4 T^5}{2} \left( \mathfrak{z}_3(w^2(1) - 1) - \mathfrak{z}_1^2(w(1)\mathfrak{z}_1 - \mathfrak{z}_2) + 0_{\text{a.s.}}(T) \right), \\ N(T; \dot{X}, X) &\stackrel{d}{=} \frac{\sigma^4 T^4}{4} \left( 4\mathfrak{z}_2(w(1)\mathfrak{z}_1 - \mathfrak{z}_2) - \mathfrak{z}_1^2(w^2(1) - 1) + 0_{\text{a.s.}}(T) \right), \end{aligned}$$

leading to (3.7). If  $X(0) = \dot{X}(0) = 0$ , then equalities in (3.7) hold for every  $T > 0$ .

IV(A). COMPLEX ROOTS:  $p = \sqrt{-1}\nu$ ,  $\nu > 0$ . This case is the subject of [13]. Below we outline the main steps.

By (3.9),

$$\begin{aligned} X(t) &= X(0) \cos(\nu t) + \frac{\dot{X}(0)}{\nu} \sin(\nu t) + \frac{\sigma}{\nu} \int_0^t \sin(\nu(t-s)) dW(s), \\ \dot{X}(t) &= -X(0)\nu \sin(\nu t) + \dot{X}(0) \cos(\nu t) + \sigma \int_0^t \cos(\nu(t-s)) dW(s). \end{aligned}$$

By Proposition 4.1, as  $T \rightarrow \infty$ , the pair

$$\left( \frac{\sqrt{2}}{\sqrt{T}} \int_0^{Tt} \sin(\nu s) dW(s), \frac{\sqrt{2}}{\sqrt{T}} \int_0^{Tt} \cos(\nu s) dW(s), t \in [0, 1] \right)$$

converges in distribution to a two-dimensional standard Brownian motion  $(w_1, w_2)$ . Then

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T^2} \int_0^T X^2(t) dt &\stackrel{d}{=} \frac{\sigma^2}{2\nu^2} \int_0^1 (w_1^2(s) + w_2^2(s)) ds, \\ \lim_{T \rightarrow \infty} \frac{1}{T^2} \int_0^T \dot{X}^2(t) dt &\stackrel{d}{=} \frac{\sigma^2}{2} \int_0^1 (w_1^2(s) + w_2^2(s)) ds, \\ \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T X(t) dW(t) &\stackrel{d}{=} \frac{\sigma}{2\nu} \int_0^1 (w_1(s) dw_2(s) - w_2(s) dw_1(s)), \\ \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \dot{X}(t) dW(t) &\stackrel{d}{=} \frac{\sigma}{2} \int_0^1 (w_1(s) dw_1(s) + w_2(s) dw_2(s)). \end{aligned}$$

Finally, (4.13) implies  $X^2(T) = T^2 0_{\text{a.s.}}(T)$ . Then both equalities in (3.8) follow from (4.1). For details, see [13, Sections 2–4].

IV(B). COMPLEX ROOTS:  $p = \lambda + \sqrt{-1}\nu$ ,  $\lambda > 0$ ,  $\nu > 0$ . The fundamental system of solutions of (3.10) in this case is

$$x_1(t) = \frac{e^{\lambda t}}{\nu} (\nu \cos \nu t - \lambda \sin \nu t), \quad x_2(t) = \frac{1}{\nu} e^{\lambda t} \sin \nu t.$$

Therefore,

$$(5.14) \quad X(t) = \frac{e^{\lambda t}}{\nu} \left( (\dot{X}(0) - X(0)\lambda) \sin \nu t + X(0)\nu \cos \nu t \right) + \frac{\sigma}{\nu} \int_0^t e^{\lambda(t-s)} \sin \nu(t-s) dW(s),$$

$$(5.15) \quad \dot{X}(t) = \lambda X(t) + \nu Y(t),$$

$$(5.16) \quad Y(t) = \frac{e^{\lambda t}}{\nu} \left( (\dot{X}(0) - X(0)\lambda) \cos \nu t - X(0)\nu \sin \nu t \right) + \frac{\sigma}{\nu} \int_0^t e^{\lambda(t-s)} \cos \nu(t-s) dW(s).$$

Define Gaussian random variables

$$(5.17) \quad \mathbf{u}_c = \frac{\dot{X}(0) - X(0)\lambda}{\nu} + \frac{\sigma}{\nu} \int_0^{+\infty} e^{-\lambda t} \cos \nu t dW(t),$$

$$\mathbf{u}_s = -X(0) + \frac{\sigma}{\nu} \int_0^{+\infty} e^{-\lambda t} \sin \nu t dW(t)$$

and the functions

$$V_c(t) = e^{\lambda t} \cos \nu t, \quad V_s(t) = e^{\lambda t} \sin \nu t.$$

Then

$$(5.18) \quad X(t) = \mathbf{u}_c V_s(t) - \mathbf{u}_s V_c(t) + e^{\lambda t} 0_{\text{a.s.}}(t), \quad Y(t) = \mathbf{u}_s V_s(t) + \mathbf{u}_c V_c(t) + e^{\lambda t} 0_{\text{a.s.}}(t).$$

By (4.5),

$$(5.19) \quad D(T; X, \dot{X}) = \frac{e^{4\lambda T} \nu^4}{16\lambda^2(\lambda^2 + \nu^2)} \left( (\mathbf{u}_c^2 + \mathbf{u}_s^2)^2 + 0_{\text{a.s.}}(T) \right).$$

Next, define Gaussian random variables

$$(5.20) \quad \mathbf{h}_c(T) = \int_0^T e^{\lambda(t-T)} \cos \nu t dW(t), \quad \mathbf{h}_s(T) = \int_0^T e^{\lambda(t-T)} \sin \nu t dW(t).$$

For each  $T > 0$ , the vector  $(\mathbf{h}_c(T), \mathbf{h}_s(T))$  is bivariate normal with zero mean,

$$(5.21) \quad \sigma_c^2(T) = \mathbb{E} \mathbf{h}_c^2(T) = \frac{1}{4\lambda} \left( 1 + \frac{\lambda}{\sqrt{\lambda^2 + \nu^2}} \cos(2\nu T - \phi) \right) + o(1),$$

$$(5.22) \quad \sigma_s^2(T) = \mathbb{E} \mathbf{h}_s^2(T) = \frac{1}{4\lambda} \left( 1 - \frac{\lambda}{\sqrt{\lambda^2 + \nu^2}} \cos(2\nu T - \phi) \right) + o(1),$$

$$(5.23) \quad \sigma_{cs}(T) = \mathbb{E} \mathbf{h}_c(T) \mathbf{h}_s(T) = \frac{1}{4\sqrt{\lambda^2 + \nu^2}} \sin(2\nu T - \phi) + o(1),$$

where  $\cos \phi = \lambda/\sqrt{\lambda^2 + \nu^2}$ ,  $\sin \phi = \nu/\sqrt{\lambda^2 + \nu^2}$ , and  $o(1)$  denotes a non-random function  $\epsilon = \epsilon(T)$  such that  $\lim_{T \rightarrow \infty} \epsilon(T) = 0$ . Thus, the family  $\{\mathbf{h}_c(T), \mathbf{h}_s(T), T > 0\}$  is relatively compact, with limit points being bivariate Gaussian vectors  $(\mathbf{h}_c, \mathbf{h}_s)$ . Computations similar to (4.10) show that each vector  $(\mathbf{h}_c, \mathbf{h}_s)$  is independent of  $(\mathbf{u}_c, \mathbf{u}_s)$ .

On the other hand,

$$(5.24) \quad \begin{aligned} N(T; V_c, V_s) &= \sigma e^{3\lambda T} (\sigma_c^2(T) \mathfrak{h}_s(T) - \sigma_{cs} \mathfrak{h}_c(T)), \\ N(T; V_s, V_c) &= \sigma e^{3\lambda T} (\sigma_s^2(T) \mathfrak{h}_c(T) - \sigma_{cs} \mathfrak{h}_s(T)). \end{aligned}$$

To complete the proof, it remains to express  $N(T; X, \dot{X})$  and  $N(T; \dot{X}, X)$  in terms of  $N(T; V_c, V_s)$  and  $N(T; V_s, V_c)$  using (4.6) and Proposition 4.2.

*This completes the proof of Theorem 3.2.*

**Proof of Theorem 3.3.** The proof is straightforward analysis of the computations in the proof of Theorem 3.2, with the goal to find suitable normalization leading to the Gaussian limit. Let us illustrate this in the most interesting case, when  $p = \lambda + \sqrt{-1}\nu$  and the corresponding rate matrix  $R_T$  is not diagonal. In this case, (5.14)–(5.18) and (5.20), together with Proposition 4.2, imply

$$(5.25) \quad \begin{aligned} \begin{pmatrix} \int_0^T X(s) dW(s) \\ \int_0^T \dot{X}(s) dW(s) \end{pmatrix} &= -\frac{e^{\lambda T}}{\nu} \begin{pmatrix} -1 & 0 \\ -\lambda & \nu \end{pmatrix} \begin{pmatrix} \mathbf{u}_s & -\mathbf{u}_c \\ \mathbf{u}_c & \mathbf{u}_s \end{pmatrix} \begin{pmatrix} \mathfrak{h}_c(T) + 0_{\text{a.s.}}(T) \\ \mathfrak{h}_s(T) + 0_{\text{a.s.}}(T) \end{pmatrix} \\ &= A_T^{-1} (B(\mathbf{u}_s, \mathbf{u}_c))^{-1} \begin{pmatrix} \mathfrak{h}_c(T) + 0_{\text{a.s.}}(T) \\ \mathfrak{h}_s(T) + 0_{\text{a.s.}}(T) \end{pmatrix}. \end{aligned}$$

On the other hand, according to (4.1),

$$\begin{pmatrix} \widehat{\theta}_{2,T} - \theta_2 \\ \widehat{\theta}_{1,T} - \theta_1 \end{pmatrix} = \Psi_T^{-1} \begin{pmatrix} \int_0^T X(s) dW(s) \\ \int_0^T \dot{X}(s) dW(s) \end{pmatrix}.$$

That is,

$$B(\mathbf{u}_s, \mathbf{u}_c) A_T \Psi_T (\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}) = \begin{pmatrix} \mathfrak{h}_c(T) + 0_{\text{a.s.}}(T) \\ \mathfrak{h}_s(T) + 0_{\text{a.s.}}(T) \end{pmatrix}.$$

*This completes the proof of Theorem 3.3.*

**Proof of Theorem 3.4.** The proof is straightforward analysis of the computations in the proof of Theorem 3.2, this time with an emphasis on the asymptotic behavior of the matrix  $\Psi_T$  and the vector  $\int_0^T \mathbf{X}(t) dW(t)$ . The analysis is easy in the ergodic case and also when  $p = 0$  or  $p = \sqrt{-1}\nu$ . When  $p = \lambda + \sqrt{-1}\nu$ , the proof is essentially complete after (5.25).

Here are the results when  $p > 0$ . If  $q < p$ , then

$$\int_0^T \mathbf{X}(t) dW(t) = e^{pT} \zeta_p (\eta_p(T) + 0_{\text{a.s.}}(T)) \mathbf{b}_p, \quad \Psi_T = \frac{\zeta_p^2}{2p} e^{2pT} (\mathbf{b}_p \mathbf{b}_p^\top + 0_{\text{a.s.}}(T)).$$

If  $q = p$ , then

$$\int_0^T \mathbf{X}(t) dW(t) = T e^{pT} \zeta (\eta_{p,2}(T) + 0_{\text{a.s.}}(T)) \mathbf{b}_p, \quad \Psi_T = \frac{\zeta^2}{2p} T^2 e^{2pT} (\mathbf{b}_p \mathbf{b}_p^\top + 0_{\text{a.s.}}(T)).$$

Note also that

$$(\mathbf{b}_p \mathbf{b}_p^\top)^2 = (1 + p^2) \mathbf{b}_p \mathbf{b}_p^\top.$$

*This completes the proof of Theorem 3.4.*



## 6. SUMMARY AND DISCUSSION

The maximum likelihood estimator (2.17) in CAR(1), that is, the one-dimensional OU process (2.16), has three types of asymptotic regimes, depending on the sign of the parameter  $\theta$ . For CAR(2), that is, the second-order equation (1.1) with two unknown parameters, while the MLE still has the same form and is strongly consistent for all values of the parameters, the number of different asymptotic regimes is nine. This jump in complexity underlines the challenges related to the analysis of the general estimation problem, either for the  $N$ -th order linear equation (CAR( $N$ )) or an  $N$ -by- $N$  system.

If equation (2.22) has real roots and one of them is positive, then the following two features seem to be common:

- (1) The rate of convergence of the estimators is determined by the smaller root of equation (2.22);
- (2) The asymptotic behavior of the normalized log-likelihood ratio is determined by the larger root of equation (2.22).

It is interesting that the problem is much easier for homogeneous equations, that is, multi-dimensional analogues of the geometric Brownian motion. In those models, estimation of the drift matrix in any number of dimensions leads to the LAN situation as long as the diffusion matrix is non-degenerate; see [9] for details.

The finite-difference equation arising from discretization of (1.1) presents other challenges. In particular, the noise sequence driving the equation is no longer independent; see [13, Section 5] for details. While these difficulties can be resolved in the ergodic case ([4]), the general case remains unsolved.

## 7. ACKNOWLEDGEMENT

The work of both authors was partially supported by the NSF Grant DMS-0803378.

## REFERENCES

- [1] M. Arató, *Linear stochastic systems with constant coefficients: a statistical approach*, Lecture Notes in Control and Information Sciences, vol. 45, Springer-Verlag, Berlin, 1982.
- [2] G. K. Basak and P. Lee, *Asymptotic properties of an estimator of the drift coefficients of multidimensional Ornstein-Uhlenbeck processes that are not necessarily stable*, Electron. J. Stat. **2** (2008), 1309–1344.
- [3] I. V. Basawa and D. J. Scott, *Asymptotic optimal inference for nonergodic models*, Lecture Notes in Statistics, vol. 17, Springer-Verlag, New York, 1983.
- [4] P. J. Brockwell, R. A. Davis, and V. Yang, *Continuous-time Gaussian autoregression*, Statistica Sinica **17** (2007), no. 1, 63–80.
- [5] P. D. Feigin, *Maximum likelihood estimation for continuous-time stochastic processes*, Advances in Appl. Probability **8** (1976), no. 4, 712–736.
- [6] P. E. Greenwood and W. Wefelmeyer, *Asymptotic minimax results for stochastic process families with critical points*, Stochastic Process. Appl. **44** (1993), no. 1, 107–116.
- [7] I. A. Ibragimov and R. Z. Khasminskii, *Statistical estimation: Asymptotic theory*, Applications of Mathematics, vol. 16, Springer, 1981.
- [8] J. Jacod and A. N. Shiryaev, *Limit theorems for stochastic processes, 2nd ed.*, Grundlehren der Mathematischen Wissenschaften, vol. 288, Springer, 2003.
- [9] A. Jankunas and R. Z. Khasminskii, *Estimation of parameters of linear homogeneous stochastic differential equations*, Stochastic Process. Appl. **72** (1997), no. 2, 205–219.

- [10] P. Jeganathan, *Some aspects of asymptotic theory with applications to time series models*, *Econometric Theory* **11** (1995), no. 5, 818–887.
- [11] Yu. A. Kutoyants, *Statistical inference for ergodic diffusion processes*, Springer, 2004.
- [12] L. Le Cam and G. L. Yang, *Asymptotics in statistics: Some basic concepts*, second ed., Springer Series in Statistics, Springer-Verlag, New York, 2000.
- [13] N. Lin and S. V. Lototsky, *Undamped harmonic oscillator driven by additive Gaussian white noise: A statistical analysis*, *Commun. Stoch. Anal.* **5** (2011), no. 1, 233–250.
- [14] R. Sh. Liptser and A. N. Shiryaev, *Theory of martingales*, Mathematics and its Applications (Soviet Series), vol. 49, Kluwer Academic Publishers, Dordrecht, 1989.
- [15] R. Sh. Liptser and A. N. Shiryaev, *Statistics of random processes, I: General theory, 2nd ed.*, Applications of Mathematics, vol. 5, Springer, 2001.
- [16] H. Luschgy, *Local asymptotic mixed normality for semimartingale experiments*, *Probab. Theory Related Fields* **92** (1992), no. 2, 151–176.

*Current address*, N. Lin: Department of Mathematics, USC, Los Angeles, CA 90089 USA, tel. (+1) 213 821 1480; fax: (+1) 213 740 2424

*E-mail address*, N. Lin: [nlin@usc.edu](mailto:nlin@usc.edu)

*Current address*, S. V. Lototsky (corresponding author): Department of Mathematics, USC, Los Angeles, CA 90089 USA, tel. (+1) 213 740 2389; fax: (+1) 213 740 2424

*E-mail address*, S. V. Lototsky: [lototsky@usc.edu](mailto:lototsky@usc.edu)

*URL*: <http://www-rcf.usc.edu/~lototsky>