HAL open science

Asymptotic near-efficiency of the "Gibbs-energy (GE) and empirical-variance" estimating functions for fitting Matérn models - II: Accounting for measurement errors via "conditional GE mean"

Didier A. Girard

▶ To cite this version:

Didier A. Girard. Asymptotic near-efficiency of the "Gibbs-energy (GE) and empirical-variance" estimating functions for fitting Matérn models - II: Accounting for measurement errors via "conditional GE mean". Statistics and Probability Letters, 2020, Statistics and Probability Letters, 162, pp.108726:1-11. 10.1016/j.spl.2020.108726 . hal-00413693v3

HAL Id: hal-00413693
https://hal.science/hal-00413693v3

Submitted on 7 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Asymptotic near-efficiency of the "Gibbs-energy (GE) and empirical-variance" estimating functions for fitting Matérn models – II: Accounting for measurement errors via "conditional GE mean" *

Didier A. Girard

CNRS and Grenoble-Alpes University, Lab. LJK, F-38000, Grenoble, France

Abstract

Consider one realization of a continuous-time Gaussian process Z which belongs to the Matérn family with known "regularity" index $\nu > 0$. For estimating the autocorrelation-range and the variance of Z from n observations on a fine grid, we studied in Girard (2016) the GE-EV method which simply retains the empirical variance (EV) and equates it to a candidate "Gibbs energy (GE)", i.e. the quadratic form $\mathbf{z}^T R^{-1}\mathbf{z}/n$ where $\mathbf{z}$ is the vector of observations and R is the autocorrelation matrix for $\mathbf{z}$ associated with a candidate range. The present study considers the case where the observation is $\mathbf{z}$ plus a Gaussian white noise whose variance is known. We propose to simply bias-correct EV and to replace GE by its conditional mean given the observation. We show that the ratio of the large-n mean squared error of the resulting CGEM-EV estimate of the range-parameter to the one of its maximum likelihood estimate, and the analog ratio for the variance-parameter, have the same behavior than in the no-noise case: they both converge, when the grid-step tends to 0, toward a constant, only function of ν, surprisingly close to 1 provided ν is not too large. We also obtain, for all ν, convergence to 1 of the analog ratio for the microergodic-parameter.

1 Introduction

We consider time-series of length n obtained by observing, at n equispaced times, a continuous-time process Z which is Gaussian, has mean zero and an autocorrelation function which belongs to the Matérn family with "regularity" index $\nu > 0$. See the Introduction of Girard (2016) and the references therein for comments on this popular family. We just recall, for notational completeness, that a Matérn processes on $\mathbb{R}$ can be specified by its spectral density over $(-\infty, +\infty)$ where θ designates the so-called "(inverse) range parameter":

$$f^*_{\nu,b,\theta}(\omega) = \tau^2\, g^*_{\nu,\theta}(\omega), \quad \text{with} \quad g^*_{\nu,\theta}(\omega) := \frac{C_\nu\, \theta^{2\nu}}{(\theta^2+\omega^2)^{\nu+\frac{1}{2}}} \quad \text{where} \quad C_\nu = \frac{\Gamma\left(\nu+\frac{1}{2}\right)}{\sqrt{\pi}\Gamma(\nu)}. \tag{1.1}$$

*The previous version (version 2) considered both the case with measurement errors (also called "nugget-effect" or simply "noise") and the no-noise case. The no-noise case is now in Girard (2016) with more detailed proofs and two additional (w.r.t. version 2) results: a consistency result is proved and the restriction $\nu \geq 1/2$ is eliminated. This version 3 is devoted to the case with measurement errors, and also gives the analogs of these two additional results.

In this paper τ^2 is the variance of $Z(t)$ (it is easily checked that $\int_{-\infty}^{\infty} g^*_{\nu,\theta}(\omega)d\omega = 1$).

As in Girard (2016), we are concerned here with "dense" grid for the observation times (or "locations") in the sense that the sampling period $\delta > 0$ is "small" enough. Stein (1999, Chapter 3) shows that a standard (i.e. fixed $\delta > 0$) large-n asymptotic analysis followed by a less standard small-δ analysis yields useful theoretical insights. This is precisely the asymptotic framework we use here.

But, we assume now that there are Gaussian i.i.d. measurement errors, or, equivalently for the parametric inference point of view we take here, there is a geostatistical "nugget effect", with known variance σ_{N}^2. And we assume that ν is known. That is, given known $\nu > 0, \delta > 0$ and $\sigma_{\mathrm{N}} > 0$, one observes only a vector of size n which, after scaling by σ_{N}, has a distribution satisfying the model:

$$\frac{\mathbf{y}}{\sigma_{\mathrm{N}}} \sim N(0, b_0 R_{\theta_0} + I_n) \text{ where } b_0 = \frac{\tau_0^2}{\sigma_{\mathrm{N}}^2} \tag{1.2}$$

with I_n denoting the identity matrix and R_θ the Toeplitz matrix of coefficients $[R_\theta]_{j,k} = K_{\nu,\theta}(\delta|j-k|)$, $j,k = 1,\cdots,n$, with $K_{\nu,\theta}(t) = \int_{-\infty}^{\infty} g^*_{\nu,\theta}(\omega)e^{i\omega t}d\omega$ (see e.g. Stein (1999, Section 2.5) for expressions for these autocorrelation functions $K_{\nu,\theta}(\cdot)$). We can thus call b_0 the true signal-to-noise ratio (SNR). Notice that one may already expect that the results of our present study for the particular case $b_0 \gg 1$ and, say, $\sigma_{\mathrm{N}}^2 = 1$, will approximately coincide with those of the "no-noise" situation of Girard (2016) (where b_0 designated the true variance of Z).

The CGEM-EV method, introduced in the first arXiv version of Girard (2012) and that we study here, is an extension of GE-EV (which was studied in Girard (2016)) to the case of noisy observations (or nugget-effect) of known variance, that we consider as a "natural" extension. Indeed, recalling that, firstly GE-EV reverses the roles played by the variance and the range-parameter in the well known hybrid method of Zhang and Zimmerman (2007) (where a "rough" estimate of the range is used) and uses the "rough" empirical variance, it seems natural to merely correct this naive, yet near-efficient (in the sense stated by Girard, 2016), estimator of τ_0^2, by its known bias. Thus we define

$$\hat{\tau}^2_{\mathrm{EV}|\sigma_{\mathrm{N}}} := \frac{1}{n}\mathbf{y}^T\mathbf{y} - \sigma_{\mathrm{N}}^2 \quad \text{and} \quad \hat{b}_{\mathrm{EV}|\sigma_{\mathrm{N}}} := \frac{\hat{\tau}^2_{\mathrm{EV}|\sigma_{\mathrm{N}}}}{\sigma_{\mathrm{N}}^2}. \tag{1.3}$$

The second ingredient of CGEM-EV consists of replacing the maximization of the likelihood (ML) w.r.t. θ by the following estimating equation in θ: denoting by $A_{b,\theta}$ the "signal extraction" matrix (see Section 2)

$$A_{b,\theta} := bR_\theta(I_n + bR_\theta)^{-1},$$

find, with b fixed at $\hat{b}_{\mathrm{EV}|\sigma_{\mathrm{N}}}$ a root θ of

$$\mathrm{CGEM}(b,\theta) - b\sigma_{\mathrm{N}}^2 \quad \text{with } \mathrm{CGEM}(b,\theta) := \frac{1}{n}\mathbf{y}^T A_{b,\theta} {R_\theta}^{-1} A_{b,\theta}\mathbf{y} + \left(\frac{b}{n}\mathrm{tr}(I_n - A_{b,\theta})\right)\sigma_{\mathrm{N}}^2. \tag{1.4}$$

Recall that, if the un-noisy discretely sampled process, say $\mathbf{z}$, were observed, the second ingredient of the GE-EV method (the equation which replaces (1.4)) would consist of finding the matching between the variance τ^2 and $n^{-1}\mathbf{z}^T {R_\theta}^{-1}\mathbf{z}$, a quantity we call the candidate "Gibbs energy" (GE, in short) of $\mathbf{z}$. On the other hand, by classic manipulations, one can check that $\mathrm{CGEM}(b,\theta)$ is the conditional mean of this GE given $\mathbf{y}, \sigma_{\mathrm{N}}$ and the candidate (b,θ). Let us now combine a well known result about the use of likelihood scores in case of incomplete data (e.g. Heyde, Section 7.4.1), and the remark recalled in Girard (2016) that $n^{-1}\mathbf{z}^T {R_\theta}^{-1}\mathbf{z} - \tau^2$ is, up to a strictly positive deterministic factor, the derivative of the log-likelihood of $\mathbf{z}$ w.r.t. b. We thus deduce that the proposal (1.4) is, in fact, (and still up to a (> 0) factor) the *likelihood score w.r.t.* b (and *not* θ !)

when only $\mathbf{y}$ is observed. Thus the first heuristic justification in Girard (2016) could be repeated here, except that the analog of the constrained ML b-estimator function (i.e., in the nototation of Girard (2016), $\theta \rightsquigarrow \hat{b}_{\mathrm{ML}}(\theta)$; recall that the heuristic given there, was that adjusting θ so that this function be matched to a rough estimate b_1 for the variance is a useful idea, at least in the infill framework) is no more explicit, and the score equation may be unsufficiant to define such a function (note also that the theoretical result of Kaufman and Shaby (2013) only deals with the no-noise case).

In the following, we denote by $\hat{\theta}_{\mathrm{GEV}|\sigma_{\mathrm{N}}}$ this range-parameter estimate (in practice, for a reason suggested at the third paragraph of Section 4, we chose the smallest root in case of multiple roots).

Note that, b being fixed at $\hat{b}_{\mathrm{EV}|\sigma_{\mathrm{N}}}$, computing CGEM$(b, \theta)$ at candidate θ does not require to apply ${R_\theta}^{-1}$ (since, obviously $A_{b,\theta}{R_\theta}^{-1}A_{b,\theta} = bR_\theta\left(I_n + bR_\theta\right)^{-1}$) and it is thus the condition number of $R_\theta + b^{-1}I_n$, not of R_θ (as it was the case for GE-EV), which controls the numerical stability of the computation. Numerical experiments by Lim et al. (2017) provide a detailed analysis of this condition-number for the Matérn covariance. Thus, as it was already known from experiments in kriging or ML computations, numerical instability can be alleviated by adding in the model an, even small, nugget effect. That is, even for un-noisy observations, it may be useful (and sometimes mandatory) to use CGEM-EV, instead of GE-EV, with a small a priori fixed σ_{N}^2 (the impact of such a prior value is studied in the last experiment of Girard (2017)). Let us add that in the latter paper, other comments are given (especially in its Secttion 2.3) about computational aspects of CGEM-EV as an alternative to ML when σ_{N}^2 is known.

In this article, we shall provide an asymptotic justification for CGEM-EV, as compared to ML, identical to that already obtained for GE-EV in the no-noise case, except we do not give a precise meaning of the "small-ness" of δ which is sufficient for guaranteeing an asymptotic consistency of $\hat{\theta}_{\mathrm{GEV}|\sigma_{\mathrm{N}}}$. Recall that the "$\nu$ not too-far from $1/2$" condition is required to obtain appealing near-efficiency results (more precisely, e.g., $0 < \nu \leq 3$ implies a mean-squared-error inefficiency less than 1.33 in the asymptotic framework we use). In practice, $\nu > 3$ is rarely used, see e.g. Stein (1999), Gaetan and Guyon (2010)).

We hope that the theoretical justification obtained here can be extended to more computationally complex settings. Indeed, this approach is clearly not limited to observations on a one dimensional lattice, and is potentially not limited to regular grids (a weighted version, with Riemann-sum type coefficients, of the empirical variance may then be useful). Successful experiments with CGEM-EV and its Riemann-sum version, with various simulated two-dimensional Matérn random fields, are described in Girard (2017). See also the Mathematica Demo (Girard 2014) we produced so that any one can easily assess CGEM-EV for the case $\nu = 1/2$.

The rest of this article is structured exactly as Girard (2016), except that, in addition, the infill framework is somewhat discussed at the end of Section 4.

2 Further notations and some properties of the spectral densities for Matérn time-series

Let us recall that, as in Girard (2016), we choose the vocabulary here (i.e. "time" in place of "space") since we use in numerous places of the paper the now classical time-series theory. Set-up (1.2) is equivalent to assuming that only a Gaussian time-series Z_δ, defined by $Z_\delta(i) := Z(\delta i)$, perturbed by a Gaussian white noise, independant of Z, is observed at $i = 1, 2, \cdots, n$. From the well known aliasing formula (e.g. Section 3.6 of Stein (1999)), the spectral density on $(-\pi, \pi]$ of the observed series is

$$f^{\delta}_{\nu,b,\theta,\sigma_{\rm N}} = \sigma_{\rm N}^2 \left(b\, g^{\delta}_{\nu,\theta} + \frac{1}{2\pi} \right) \quad \text{with} \quad g^{\delta}_{\nu,\theta}(\cdot) := \frac{1}{\delta} \sum_{k=-\infty}^{\infty} g^{*}_{\nu,\theta} \left(\frac{\cdot + 2k\pi}{\delta} \right). \tag{2.1}$$

Recall that, when $\nu-1/2$ is an integer, then $g^{\delta}_{\nu,\theta}$ coincide with particular ARMA spectral densities with a *constrained* vector of parameters.

In order to simplify the statement of the results here (and their proofs), it is convenient to introduce the following weight function $a^{\delta}_{b,\theta}(\cdot)$ over $(-\pi, \pi]$, that we call the candidate filter for given (b, θ)

$$a^{\delta}_{b,\theta}(\cdot) := g^{\delta}_{\nu,\theta}(\cdot) / \left(g^{\delta}_{\nu,\theta}(\cdot) + (2\pi b)^{-1} \right). \tag{2.2}$$

Indeed, as is well known from the signal extraction literature, $a^{\delta}_{b,\theta}$ is the frequency response of the "optimal (if b, θ were the true parameters)" convolution of the perturbed series if it were observed over $\mathbb{Z}$; see e.g. Section 4.11 of Shumway and Stoffer 2006 for details, and Girard (2012) also for related well known "best extracting" properties of applying the matrix $A_{b,\theta}$.

A function which will play an important role in this article (as it was the case in Girard (2016)) is the derivative of $\log(g^{\delta}_{\nu,\theta}(\cdot))$ w.r.t. θ; we just recall that it has the following useful expression:

$$h^{\delta}_{\nu,\theta} = \frac{2\nu}{\theta} \left(1 - \frac{g^{\delta}_{\nu+1,\theta}}{g^{\delta}_{\nu,\theta}} \right), \quad \text{where} \quad h^{\delta}_{\nu,\theta} := \partial \log(g^{\delta}_{\nu,\theta}) / \partial\theta. \tag{2.3}$$

For any $f : [-\pi, \pi] \to \mathbb{R}$, s.t. $\int_{-\pi}^{\pi} w(\lambda) f(\lambda) \mathrm{d}\lambda \neq 0$, where $w(\cdot) > 0$ is a weight function (we, in fact, only use $w := [a^{\delta}_{b,\theta}]^2$) we define the weighted coefficient of variation of f by

$$J_w(f) := \left\{ \frac{1}{\int w} \int w \left| f - \frac{1}{\int w} \int wf \right|^2 \right\} \Big/ \left(\frac{1}{\int w} \int wf \right)^2 = \frac{\frac{1}{\int w} \int wf^2}{\left(\frac{1}{\int w} \int wf \right)^2} - 1\ . \tag{2.4}$$

Above and throughout this paper, "$\int$" will denote integrals over $[-\pi, \pi]$. Omitting the indexes δ and ν, we will also use the notation g_0 (resp. h_0) for the function $g^{\delta}_{\nu,\theta}$ (resp. $h^{\delta}_{\nu,\theta}$) when $\theta = \theta_0$. B (resp. Θ) will denote any compact interval not containing 0 and such that b_0 (resp. θ_0) is in the interior of B (resp. Θ) .

We now collect in the following lemmas (whose proof are postponed to an Appendix) small-δ equivalences which will be used to prove the results of the following Sections; they might be of interest also for other studies of the Matérn time-series plus white noise model:

Lemma 2.1. *For any $b > 0, \theta > 0, \nu > 0$ and $k \in \{1, 2\}$, we have as $\delta \downarrow 0$:*

$$\int \left(\frac{g^{\delta}_{\nu,\theta}}{a^{\delta}_{b,\theta}} \right)^2 \sim \int \left(g^{\delta}_{\nu,\theta} \right)^2 \sim \frac{c_{1,\nu}}{\delta\theta}, \qquad \int [a^{\delta}_{b,\theta}]^2 \left(\frac{g^{\delta}_{\nu+1,\theta}}{g^{\delta}_{\nu,\theta}} \right)^k \sim \int \left(\frac{g^{\delta}_{\nu+1,\theta}}{g^{\delta}_{\nu,\theta}} \right)^k \sim 2\pi\, {c_{2,\nu}}^k\, \delta\theta$$

where the constants $c_{1,\nu}, c_{2,\nu}$ are given in Lemma 2.1 of Girard (2016).

Lemma 2.2. *For any $b > 0, \theta > 0, \nu > 0, k \in \{1, 2\}$ and with C_ν defined in (1.1), we have as $\delta \downarrow 0$:*

$$\int \left[a^{\delta}_{b,\theta}(\lambda) \right]^k \mathrm{d}\lambda \sim 2 \delta^{\frac{2\nu}{2\nu+1}} (2\pi C_\nu c)^{\frac{1}{2\nu+1}} \Gamma \left(k - \frac{1}{2\nu+1} \right) \Gamma \left(1 + \frac{1}{2\nu+1} \right) \quad \text{where} \quad c = b\theta^{2\nu}.$$

Now from the fact that $\delta^{\frac{2\nu}{2\nu+1}}$ dominates δ for any $\nu > 0$, and the expression (2.3) of $h^{\delta}_{\nu,\theta}$, the following corollary is easily obtained (by proving the third stated equivalence before the first one):

Corollary 2.3. *For any $b > 0, \theta > 0, \nu > 0$ and $k \in \{1, 2\}$, we have as $\delta \downarrow 0$, for the weight function $w = [a^{\delta}_{b,\theta}]^2$:*

$$J_w\left(h^{\delta}_{\nu,\theta}\right) \sim \frac{1}{\int [a^{\delta}_{b,\theta}]^2} \int \left(\frac{g^{\delta}_{\nu+1,\theta}}{g^{\delta}_{\nu,\theta}}\right)^2, \quad J_w\left(\frac{g^{\delta}_{\nu,\theta}}{[a^{\delta}_{b,\theta}]^2}\right) \sim \int [a^{\delta}_{b,\theta}]^2 \int \left(g^{\delta}_{\nu,\theta}\right)^2$$

and

$$\int [a^{\delta}_{b,\theta}]^2 \left(h^{\delta}_{\nu,\theta}\right)^k \sim \left(\frac{2\nu}{\theta}\right)^k \int [a^{\delta}_{b,\theta}]^2.$$

3 Consistency

Of course, at fixed δ, $\hat{b}_{\mathrm{EV}|\sigma_{\mathrm{N}}}$ is a consistent estimator of b_0 (see also (4.1)). We first state asymptotic properties of the (normalized) estimating equation $\mathrm{CGEM}(b, \theta) - b\sigma^2_{\mathrm{N}} = 0$ and its partial derivatives, in particular for δ "small", whose proof only requires classical techniques and the third equivalence of Corollary 2.3 (see the comments below):

Theorem 3.1. *1) We have the following three convergences in probability, uniform over $B \times \Theta$, as $n \to \infty$:*

$$\begin{aligned} \frac{\sigma_{\mathrm{N}}^{-2}\mathrm{CGEM}(b,\theta) - b}{(n^{-1}\mathrm{tr}A_{b,\theta})\, b} &= \left(\frac{\mathbf{y}^T A_{b,\theta}(I - A_{b,\theta})\mathbf{y}}{\sigma^2_{\mathrm{N}}\mathrm{tr}A_{b,\theta}} - 1\right) \\ &\to \frac{1}{\int_{-\pi}^{\pi} a^{\delta}_{b,\theta}(\lambda)} \int_{-\pi}^{\pi} [a^{\delta}_{b,\theta}(\lambda)]^2 \left(\frac{b_0 g^{\delta}_{\nu,\theta_0}(\lambda)}{b g^{\delta}_{\nu,\theta}(\lambda)} - 1\right) d\lambda \quad =: \phi(\delta, b, \theta, b_0, \theta_0), \text{ say,} \end{aligned}$$

$$\frac{\partial}{\partial b}\left(\frac{\mathrm{CGEM}(b,\theta)}{\sigma^2_{\mathrm{N}}} - b\right) \to \frac{-1}{2\pi}\left(\int_{-\pi}^{\pi} [a^{\delta}_{b,\theta}(\lambda)]^2 d\lambda - 2\int_{-\pi}^{\pi} \left([a^{\delta}_{b,\theta}(\lambda)]^2 - [a^{\delta}_{b,\theta}(\lambda)]^3\right)\left(\frac{b_0 g_0(\lambda)}{b g^{\delta}_{\nu,\theta}(\lambda)} - 1\right) d\lambda\right),$$

$$\frac{\partial}{\partial\theta}\frac{\mathrm{CGEM}(b,\theta)}{\sigma^2_{\mathrm{N}}} \to \frac{-b}{2\pi}\left(\int_{-\pi}^{\pi} [a^{\delta}_{b,\theta}(\lambda)]^2 h^{\delta}_{\nu,\theta}(\lambda) d\lambda + \int_{-\pi}^{\pi} \left(2[a^{\delta}_{b,\theta}(\lambda)]^3 - [a^{\delta}_{b,\theta}(\lambda)]^2\right) h^{\delta}_{\nu,\theta}(\lambda)\left(\frac{b_0 g_0(\lambda)}{b g^{\delta}_{\nu,\theta}(\lambda)} - 1\right) d\lambda\right).$$

2) When $\delta \downarrow 0$, we have

$$\phi(\delta, b, \theta, b_0, \theta_0) \to 2\nu(2\nu+1)^{-1}\left(\frac{b_0\theta_0^{2\nu}}{b\theta^{2\nu}} - 1\right).$$

3) There exists a strictly positive function $\bar{\delta}(\nu, b_0, \theta_0)$ such that $0 < \delta \leq \bar{\delta}(\nu, b_0, \theta_0)$ implies that the large-n limit in probability of $\frac{\partial}{\partial\theta}\mathrm{CGEM}(b, \theta)$ evaluated at (b_0, θ_0) is stricly negative.

As it was the case for Part 1 of Theorem 3.1 of Girard (2016), the first part here is in fact not restricted to the Matérn family. Indeed, it only requires regularity conditions on $g^{\delta}_{\nu,\theta}(\cdot)$, and its strict positivity, which are well fulfilled; and the three limits of 1) can directly be obtained, albeit more tediously than in the no-noise case, from classical large-n theoretical results about quadratic forms constructed from a product of powers, possibly negative, of Toeplitz matrices (e.g. Azencott and Dacunha-Castelle (1986)).

The second part of Theorem 3.1 is, on the contrary, a consequence of specific properties of the Matérn family, and, in fact, it can be proved by the same techniques as those used in Section 3 of Stein (1999). Let us comment this small–δ equivalent associated with the first p-limit of 1). Firstly, by examining the analog previous results in the no-noise case, we see that the first of these previous results is well a "particular case" of the first limit above by setting $a^{\delta}_{b,\theta}$ to 1, which is well in agreement with the guess that the no-noise case corresponds to $a^{\delta}_{b,\theta} = 1$ (notice that a similar remark can be made for the terms of the Jacobian given in Proof of Part 1 of Theorem 4.1 of Girard (2016) which are seen as particular case of the second and third limits above with $a^{\delta}_{b,\theta}$ set to 1). Secondly, always compared to the no-noise case, the small-δ limit of the p-limit (after normalization) of the equation $\sigma_{\mathrm{N}}^{-2}\mathrm{CGEM}(b, \theta) = b$ to be solved in θ, is unchanged except for a constant factor (in fact this factor could have been eliminated if we had normalized by $(n^{-1}\mathrm{tr}A^2_{b,\theta})b$ in place of $(n^{-1}\mathrm{tr}A_{b,\theta})b$, but this is unimportant and it seems more natural to choose $\mathrm{tr}A_{b,\theta}$ since it yields a simple expression for the left-hand term of the first result in Theorem 3.1 (expression given in parentheses)).

Let us thus recall that, if b is fixed at any value b_1, then the unique root θ_1 of this small-δ-large-n equivalent equation will satisfy $b_1\theta_1^{2\nu} = b_0\theta_0^{2\nu}$. This indeed gives some support to the extension to CGEM-EV of the first heuristic for GE-EV in Girard (2016), as is discussed in the Introduction.

As to the third part of Theorem 3.1, the existence of such a function $\bar{\delta}(\nu, b_0, \theta_0)$ is of course a consequence of the third equivalence of Corollary 2.3 and the strict positivity of $a^{\delta}_{b,\theta}$ (consequence of its definition), since, from the third result of Part 1, the limit in probability of $\frac{\partial}{\partial\theta}\mathrm{CGEM}(b, \theta)$ evaluated at (b_0, θ_0) clearly reduces to $-b_0\sigma_{\mathrm{N}}^2(2\pi)^{-1}\int a_0^2 h_0$ (indeed the second integral vanishes).

Remark 3.2. Since we do not give an explicit form for the upper-bound $\bar{\delta}(\nu, b_0, \theta_0)$, this is not a result as strong as the analog Part 3 in Girard (2016). Anyway, we believe that even the result in the no-noise case could be improved and we conjecture that the "local well-posedness" of the estimating-equation around θ_0 (namely a garantee that this derivative at (b_0, θ_0) converges in probability toward a non-zero value, as $n \to \infty$) does not requires that δ be small.

A Cramer-type consistency can now be proved (as detailed in the Appendix) by using Kessler et al. (2012) (where a survey of general asymptotic results for estimating equations is given, see their Section 1.10); precisely:

Theorem 3.3. *Assume that δ is not greater than $\bar{\delta}(\nu, b_0, \theta_0)$, then there exists a sequence of roots $\hat{\theta}_{\mathrm{GEV}|\sigma_{\mathrm{N}}}$ of the CGEM-EV equation (i.e. (1.4) with b fixed at $\hat{b}_{\mathrm{EV}|\sigma_{\mathrm{N}}}$), as n increases, which converges in probability to θ_0.*

4 Mean squared error inefficiencies of CGEM-EV to ML for the variance, range and microergodic parameters

As is common in classical (in the sense that the sampling period δ is fixed) time-series theory, the term "asymptotic variance of an estimator", denoted avar$(\cdot)$, will designate in this paper the variance of the limiting distribution of $\sqrt{n}$ times the error of this estimator; the large-n

mean squared error (MSE) of this estimator will refer to n^{-1} times its asymptotic variance. As noticed in Girard (2016), we could consider a size of $\lfloor n/\delta \rfloor$ for the n-th data set : this would only multiply all the asymptotic variances by δ and the following near-efficiency statements would be inchanged (see e.g. Brockwell et al. (2007)).

Consider first a simplified setting: the case where the microergodic parameter $c_0 = b_0\theta_0^{2\nu}$ is assumed to be known. (Note that it might be more natural to call "microergodic parameter" the product $\tau_0^2\theta_0^{2\nu}$ since one may prefer that this parameter does not change with σ_{N}; however since σ_{N} is assumed known, choosing between these two definitions will have no impact on the properties of considered estimators, identical up a known factor).

This assumption of a known c_0 is of course restrictive and the following Theorem 4.0 may be thought of as one of weak practical interest. However it is known that "reasonably accurate", even if not fully efficient, estimates of c_0 can be computed by less expensive approaches than ML in numerous contexts, and one could thus condition the model with such a "reasonable" value of c_0 plugged-in. Recall that one of these possibly reasonable approaches is to fix θ at a prior choice θ_1, and to maximize the likelihood only with respect to b: in certain common settings, this furnishes reasonable estimates of c_0 provided the a priori chosen range (i.e. θ_1^{-1}) is "fixed at something larger than the true value (θ_0^{-1})", as said in the Section 3.1 of Kaufman and Shaby (2013) where an empirical study well demonstrates this, in the no-noise case; and it is expected that this still hold under a noise of known variance, for which case this approach can be straightfuly extended.

In this simplified setting, one can equivalently focus either on the estimation of b_0 by the nonparametric estimate $\hat{b}_{\mathrm{EV}|\sigma_{\mathrm{N}}}$ defined by (1.4), or that of θ_0 by $\left(c_0/\hat{b}_{\mathrm{EV}|\sigma_{\mathrm{N}}}\right)^{1/(2\nu)}$. Let us choose the former since the asymptotic limiting law of $\hat{b}_{\mathrm{EV}|\sigma_{\mathrm{N}}}$ has a simple expression (see e.g. Azencott and Dacunha-Castelle (1986)), for δ fixed :

$$n^{1/2}\left(\hat{b}_{\mathrm{EV}|\sigma_{\mathrm{N}}} - b_0\right) \xrightarrow{\mathcal{D}} N\left(0, 4\pi \mathrm{v}_1\right) \quad \text{as} \quad n \to \infty, \quad \text{where} \quad \mathrm{v}_1 := b_0^2 \int a_0^{-2} g_0^2. \tag{4.1}$$

Note that the variance v_1 used in Girard (2016) does not designate the v_1 used above, but this previous v_1 is clearly a "particular case" by substituting 1 for $a_0(\cdot)$. In fact, the ratio of the present v_1 to the previous one, decreases to 1 as $b_0 \to \infty$, because, at each λ, the filter function $a_0(\lambda)$ obviously increases toward its limit 1, as $b_0 \to \infty$.

Now, by considering the spectral density model $f(b,\theta) : \lambda \rightsquigarrow {\sigma_{\mathrm{N}}}^2\left(b g_{\nu,\theta}^{\delta}(\lambda) + (2\pi)^{-1}\right)$ with $b\theta^{2\nu} = c_0$, as a function of only b, easily establishing that $\frac{\partial \log(f(c_0/\theta^{2\nu},\theta))}{\partial\theta} = a_{b,\theta}^{\delta}(h_{\nu,\theta}^{\delta} - 2\nu\theta^{-1})$ where $b = c_0/\theta^{2\nu}$ and using that $\frac{\partial (c_0/b)^{1/(2\nu)}}{\partial b} = (2\nu)^{-1}(\theta/b)$ where $\theta = (c_0/b)^{1/(2\nu)}$, the asymptotic Fisher information w.r.t. b is deduced and is seen to be > 0 (from the expression (2.3) of h_0 and the fact $a_{b,\theta}^{\delta} > 0$). Thus, by an application (similar, but easier here, to the way of establishing (4.3) below) of now classical time-series theory (e.g. Azencott and Dacunha-Castelle (1986)) one obtains for the ML maximizer over B under $b\theta^{2\nu} = c_0$, now denoted $\hat{b}_{\mathrm{ML}|c_0,\sigma_{\mathrm{N}}}$, as $n \to \infty$:

$$n^{1/2}\left(\hat{b}_{\mathrm{ML}|c_0,\sigma_{\mathrm{N}}} - b_0\right) \xrightarrow{\mathcal{D}} N\left(0, \operatorname{avar}(\hat{b}_{\mathrm{ML}|c_0,\sigma_{\mathrm{N}}})\right),$$

where

$$\operatorname{avar}(\hat{b}_{\mathrm{ML}|c_0,\sigma_{\mathrm{N}}}) := 4\pi b_0^2 \left(2\nu\theta_0^{-1}\right)^2 \left(\int a_0^2\left(h_0 - 2\nu\theta_0^{-1}\right)^2\right)^{-1}. \tag{4.2}$$

Now by using, in (4.2), the expression (2.3) of h_0, and the first and second equivalences of Lemma 2.1, one obtains:

Theorem 4.0. *The large-n MSE inefficiency of $\hat{b}_{\mathrm{EV}|\sigma_{\mathrm{N}}}$ relative to the ML estimator of the SNR b_0, when $c_0 = b_0\theta_0^{2\nu}$ is known, i.e. $I^0_{\delta,b_0,\theta_0} := 4\pi \mathrm{v}_1/\mathrm{avar}(\hat{b}_{\mathrm{ML}|\mathrm{c}_0,\sigma_{\mathrm{N}}})$, satisfies (with C_ν defined in (1.1)):*

$$I^0_{\delta,b_0,\theta_0} \to \frac{{C_{\nu+1}}^2}{C_{2\nu+1/2}C_{3/2}} = \frac{\sqrt{\pi}}{2}\left(\frac{\Gamma(\nu+3/2)}{\Gamma(\nu+1)}\right)^2 \Big/ \frac{\Gamma(2\nu+1)}{\Gamma(2\nu+1/2)} =: \mathrm{ineff}(\nu) \quad \textit{as} \quad \delta \downarrow 0.$$

It is important to notice that this definition of the constant $\mathrm{ineff}(\nu)$ coincides with the one used in Girard (2016) for the no-noise case. In the particular case $\nu = 1/2$, then $\mathrm{ineff}(1/2) = 1$. (Note there was a typographical error in Girard (2016) in the second expression of $\mathrm{ineff}(\nu)$: precisely the big fraction slash was omitted; however Table 4.1 of Girard (2016) which displayed numerical values of $\mathrm{ineff}(\nu)$ for certain values of ν is exact.) The Table 4.1 of Girard (2016) is not repeated here. We only wish to emphasize that the departure from 1 of $\mathrm{ineff}(\nu)$ as ν increases, is rather slow.

A second good news it that this inefficiency is not function of the true range θ_0 or the SNR b_0. Since it could be expected that these small-δ-large-n-inefficiencies become close to those obtained in the no-noise case only under $b_0 \gg 1$, the result that b_0 has no impact on $\mathrm{ineff}(\nu)$ may be thought as rather surprising. Recall that the asymptotic inefficiencies in Girard (2016) show the absence of any impact of θ_0 in the no-noise case; this was less surprising because this was already known in the case $\nu = 1/2$ from the efficiency result of Kessler (1997) concerning the naive empirical variance. Thus Theorem 4.0 is a neat extension of the efficiency result of Kessler (1997) to the case of measurement noise (after natural bias-correction of this empirical variance by subtracting σ_{N}^2), and a "near-extension" when ν does not depart too much from 1/2 in the sense that $\mathrm{ineff}(\nu)$ stays close to 1, whatever b_0 may be.

Now let us return to the case b_0 and θ_0 unknown. Let $(\hat{b}_{\mathrm{ML}|\sigma_{\mathrm{N}}}, \hat{\theta}_{\mathrm{ML}|\sigma_{\mathrm{N}}})$ be a maximizer of the likelihood function over $B \times \Theta$ when σ_{N} is known. One can use arguments similar to those used in Girard (2016) where the asymptotic behavior of the ML estimator was described in the no-noise case: the derivation of the asymptotic information matrix (see Theorem 4.3 of Chapter XIII of Azencott and Dacunha-Castelle (1986)) is classic, albeit more tedious; and the final expressions are relatively simple modifications, by merely adding in appropriate places the weight function a_0^2, precisely: $(\hat{b}_{\mathrm{ML}|\sigma_{\mathrm{N}}}, \hat{\theta}_{\mathrm{ML}|\sigma_{\mathrm{N}}})$ is a.s. consistent and satisfies, as $n \to \infty$:

$$n^{1/2}\left(\begin{bmatrix} \hat{b}_{\mathrm{ML}|\sigma_{\mathrm{N}}} \\ \hat{\theta}_{\mathrm{ML}|\sigma_{\mathrm{N}}} \end{bmatrix} - \begin{bmatrix} b_0 \\ \theta_0 \end{bmatrix}\right) \xrightarrow{\mathcal{D}} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, 4\pi \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}\right),$$

with

$$\begin{bmatrix} \sigma_1^2 \\ \sigma_{12} \\ \sigma_2^2 \end{bmatrix} := \left|\int a_0^2 h_0\right|^{-2} J_{a_0^2}(h_0)^{-1} \begin{bmatrix} b_0^2 \int a_0^2 h_0^2 \\ -b_0 \int a_0^2 h_0 \\ \int a_0^2 \end{bmatrix}. \tag{4.3}$$

Again note that the components of the vector $(\sigma_1^2, \sigma_{12}, \sigma_2^2)$ used in Girard (2016) and those of $(\sigma_1^2, \sigma_{12}, \sigma_2^2)$ used above, are asymptotic equivalents as $b_0 \to \infty$, because in addition to $a_0 \approx 1$, the functional $J_{a_0^2}(\cdot)$ also becomes close, for large SNR, to the simpler functional $J(\cdot)$ used in Girard (2016).

Concerning CGEM-EV, we claim:

Theorem 4.1. *Assuming $\delta < \bar{\delta}(\nu, b_0, \theta_0)$, let $\hat{\theta}_{\mathrm{GEV}|\sigma_{\mathrm{N}}}$ be a consistent root of the CGEM-EV equation (i.e. (1.4) with b fixed at $\hat{b}_{\mathrm{EV}|\sigma_{\mathrm{N}}}$ and σ_{N}^2 is the true noise variance).*

1) As $n \to \infty$

$$n^{1/2}\left(\hat\theta_{\mathrm{GEV}|\sigma_{\mathrm{N}}} - \theta_0\right) \xrightarrow{\mathcal{D}} N\left(0, 4\pi \mathrm{v}_2\right) \quad \textit{where } \mathrm{v}_2 = \left|\int a_0^2 h_0\right|^{-2} J_{a_0^2}(g_0/a_0^2)\int a_0^2\ .$$

2) The large-n MSE inefficiency of CGEM-EV to ML for b_0 (resp. for θ_0) being defined by $I^1_{\delta,b_0,\theta_0} := \mathrm{v}_1/\sigma_1^2$ (resp. $I^2_{\delta,b_0,\theta_0} := \mathrm{v}_2/\sigma_2^2 = J_{a_0^2}(g_0/a_0^2)J_{a_0^2}(h_0)$), these two inefficiencies have the following common limit (with $\mathrm{ineff}(\cdot)$ *as in Theorem 4.0):*

$$I^i_{\delta,b_0,\theta_0} \to \mathrm{ineff}(\nu) \quad \textit{as} \quad \delta \downarrow 0, \quad \textit{for} \quad i \in \{1,2\}.$$

Proof: Part 1 is proved in the Appendix. The limit of both $I^1_{\delta,b_0,\theta_0}$ and $I^2_{\delta,b_0,\theta_0}$ is directly deduced from the equivalences stated in Lemma 2.1 and Corollary 2.3.

Again, as noticed for σ_2^2 above, the v_2 used in Girard (2016) is the limit value of the v_2 defined in Part 1 above as $b_0 \to \infty$ for fixed δ.

Thus, as in the case c_0 known, the CGEM-EV estimates of b_0 and θ_0 are asymptotically nearly efficient provided ν is not too large, asymptotic full-efficiency being reached for ν close to 1/2. Notice it is rather surprising that these small-δ large-n inefficiencies are not function of the underlying θ_0 or of the underlying b_0.

The remark claimed in Girard (2016) that the knowledge of c_0 does not improve (in terms of small-δ-large-n MSE) the performance of ML estimation of θ_0 or the performance of the alternative to ML we have introduced, can be also claimed in the present setting of known error variance (this extension is still also easily checked).

Let us now consider the estimation of the microergodic parameter c_0. By the classical delta-method, one directly infer from (4.3) that the asymptotic variance of $\hat c_{\mathrm{ML}|\sigma_{\mathrm{N}}} := \hat b_{\mathrm{ML}|\sigma_{\mathrm{N}}}\hat\theta^{2\nu}_{\mathrm{ML}|\sigma_{\mathrm{N}}}$ is $4\pi c_0^2\left|\int a_0^2 h_0\right|^{-2} J_{a_0^2}(h_0)^{-1}\int \left|a_0\left(h_0 - 2\nu\theta_0^{-1}\right)\right|^2$ (note that there was a typographical error in Girard (2016) : "$\left|\int\right.$" must be replaced by "$\left.\int\right|$" for the related variance with $a_0 \equiv 1$ there). On the other hand, a similar derivation (albeit more tedious than in the no-noise case) can be done for $\hat c_{\mathrm{GEV}|\sigma_{\mathrm{N}}}$ starting from the asymptotic covariance matrix (detailed in the Appendix) of the vector $(\hat b_{\mathrm{EV}|\sigma_{\mathrm{N}}}, \hat\theta_{\mathrm{GEV}|\sigma_{\mathrm{N}}})$ and this gives v_3 below. Now one can easily deduce (still using the expression (2.3) of $h^\delta_{\nu,\theta}$, Lemma 2.1 and Corollary 2.3) that, this time, full-efficiency holds for *any* $\nu > 0$, more precisely:

Theorem 4.2. *Assuming $\delta < \bar\delta(\nu, b_0, \theta_0)$, let $\hat c_{\mathrm{GEV}|\sigma_{\mathrm{N}}} := \hat b_{\mathrm{EV}|\sigma_{\mathrm{N}}}\hat\theta^{2\nu}_{\mathrm{GEV}|\sigma_{\mathrm{N}}}$ where $\hat b_{\mathrm{EV}|\sigma_{\mathrm{N}}}$ and $\hat\theta_{\mathrm{GEV}|\sigma_{\mathrm{N}}}$ are defined as in Theorem 4.1, we have, with $c_0 = b_0\theta_0^{2\nu}$:*
1) as $n \to \infty$

$$n^{1/2}\left(\hat c_{\mathrm{GEV}|\sigma_{\mathrm{N}}} - c_0\right) \xrightarrow{\mathcal{D}} N\left(0, 4\pi \mathrm{v}_3\right)$$

where

$$\mathrm{v}_3 = \frac{c_0^2}{\int a_0^2 \left|\int a_0^2 h_0\right|^2}\left(J_{b_0,\theta_0}(g_0)\left|\int a_0^2\left(h_0 - 2\nu\theta_0^{-1}\right)\right|^2 + \left|\int a_0^2 h_0\right|^2\right).$$

2) The large-n MSE inefficiency of CGEM-EV to ML for c_0 is $I^3_{\delta,b_0,\theta_0} := 4\pi\mathrm{v}_3/\mathrm{avar}(\hat c_{\mathrm{ML}|\sigma_{\mathrm{N}}})$ and it holds that $I^3_{\delta,b_0,\theta_0} \to 1$ as $\delta \downarrow 0$; more precisely,

$$4\pi\mathrm{v}_3 \sim \mathrm{avar}(\hat c_{\mathrm{ML}|\sigma_{\mathrm{N}}}) \sim \frac{2c_0^2}{(2\pi)^{-1}\int a_0^2} \sim \frac{2\pi c_0^2(2\pi C_\nu c_0)^{\frac{-1}{2\nu+1}}}{\Gamma\left(2 - \frac{1}{2\nu+1}\right)\Gamma\left(1 + \frac{1}{2\nu+1}\right)}\delta^{\frac{-2\nu}{2\nu+1}} \quad \textit{as} \quad \delta \downarrow 0. \tag{4.4}$$

Again, the variance v_3 used in Girard (2016) is clearly obtained from the v_3 here by substituting 1 for $a_0(\cdot)$. As we remarked for the no-noise case, this full-efficiency of CGEM-EV concerning c_0, even for large ν, may again be thought of as a less surprising result than Part 2 of Theorem 4.1. Indeed this full-efficiency is suggested by the infill-asymptotic efficiency, mentionned as a heuristical partial justification in the Introduction, of our "proposed" estimator of c_0/b_1, with b fixed at "any" b_1 (we use here quotation marks, only for reminding the sobering fact that the strategy of using an arbitrarily fixed b_1 may provide poor estimates in practice, and thus we do not actually propose it).

Remark 4.3. One can make incidental remarks for the "case" $\delta = 1/n$. For the particular case $\nu = 1/2$, notice that n^{-1} times the right-hand expression of this small-δ equivalent (4.4) is identical, by setting $\delta := 1/n$, to $4\sqrt{2}{c_0}^{3/2}n^{-1/2}$, that is, well coincides with the variance, established in the infill asymptotic framework by Chen et al. (2000), of the normal approximation of the law of the ML estimator of c_0. See also Zhang and Zimmerman (2005) for a detailed rigorous study of this "reconciliation" between the two asymptotic frameworks. Naturally one can thus conjecture that a such coincidence still holds beyond the case $\nu = 1/2$, namely, that n^{-1} times the right-hand expression of (4.4), with $1/n$ substituted for δ, furnishes the infill asymptotic variance for both the ML or the CGEM-EV estimator of c_0 for any $\nu > 0$. For instance, this would furnish a variance of $(16/3)c_0^{7/4}n^{-1/4}$ for $\nu = 3/2$. Notice that this latter variance also coincides with a related variance for the integrated Brownian motion plus white noise model, which could be deduced form the Fisher information given by Theorem 2.3 of Kou (2004) (take $r = 2, s = 0$ using his notation and use that $\Gamma\left(2 - (2\nu+1)^{-1}\right)\Gamma\left(1 + (2\nu+1)^{-1}\right) = (2\nu+1)\,\mathrm{B}\left(r - (2\nu+1)^{-1}, s + (2\nu+1)^{-1}\right)$ where $\mathrm{B}(\cdot,\cdot)$ is the Beta function, see e.g. Weisstein (2019)).

Remark 4.4. Since the framework of Zhang and Zimmerman (2005) actually does not imposes that σ_{N} be known, their results compared with (4.4) (where σ_{N} is known) show that by estimating σ_{N} (which is often more easy to estimate than c_0; see Chen et al. (2000) for the meaning of such a claim in the infill framework), by the ML principe, one adds no further error, at least in the small-δ-large-n framework, to the ML estimator of c_0. Is is naturally expected that this still holds beyond the case $\nu = 1/2$. An extension of CGEM-EV to the case σ_{N} unknown, is commented in the following Discussion.

5 Discussion

CGEM-EV is thus a natural extension of GE-EV to the case with measurement errors of known variance, via bias-correction of the naive empirical variance and replacement of the unobserved GE function by the conditional GE mean function. We have proved here that identical near-efficiency results still hold, not only in the particular case $b_0 \gg 1$ for which such a similarity could be expected. One may be surprised by the fact that these efficiency results hold for any fixed SNR b_0, even small. However, one must keep in mind that these results deal only with large-n asymptotics at δ fixed, always followed by a small-δ analysis: one may guess that for b_0 too small, very large n may be required to "see" the stated asymptotic behaviors, and even increasingly large as δ decreases. Recall also that, even in the no-noise case, when δ decreases to 0, larger data sizes are required to be able to accurately approximate the actual law of any one of these estimates of θ_0 (or b_0) by its asymptotic form (indeed this is well known for ML estimates in the case $\nu = 1/2$ thoroughly studied by Zhang and Zimmerman 2005)). An asymptotic comparison of CGEM-EV to ML deserves thus a futur study also in the finite-δ case.

Concerning the problem of estimating multidimensional stationary Matérn fields observed on a lattice under i.i.d. noise, as already noticed for GE-EV in Girard (2016) in the no-noise case, the CGEM-EV approach is directly applicable, in theory, provided θ remains a scalar parameter, e.g. for isotropic autocorrelations. We refer to Girard (2017) for a rather extensive empirical comparison of CGEM-EV to ML in the two-dimensional case, with randomized-traces used instead of the exact traces of (1.4). There, it is noticed, in particular, that difficulties appear in case of "too small" SNRs and "too smooth" fields (for example $\nu \geq 3/2$) both for CGEM-EV and ML. The application of CGEM-EV can be practical for very large lattice sizes, even with missing data, as soon as computing analogues of $\left(I_n + \hat{b}_{\mathrm{EV}|\sigma_{\mathrm{N}}} R_\theta\right)^{-1} \mathbf{y}$ for candidate R_θ, can be done by fast iterative algorithms; indeed each iteration can be fast since applying R_θ to a vector can reduce to three multidimensional discrete (inverse) Fourier transforms. Thus extensions of the asymptotic results of this paper to such multidimensional fitting problems clearly deserves a detailed study. And a comparison of CGEM-EV with the classical (tapered) Whittle-likelihood maximization for such multidimensional fitting problems should be of interest.

CGEM-EV might be extended to the important case of unknown noise variance, via, at least, two simple ways which are described and experimented in Girard (2018). One of these two ways is to add, as a second simple estimating equation, the first derivative of the likelihood w.r.t. σ_{N}^2 for given b and θ equated to zero. Simulations in Girard (2018) demonstrate rather good performance of these two approaches, which thus warrants further exploration.

REFERENCES

Azencott, R., Dacunha-Castelle, D., 1986. Series of Irregular Observations: Forecasting and Model Building. Springer. New York.

Brockwell, P. J., Davis, R. A., and Yang Y., 2007. Continuous-time Gaussian autoregression. Statistica. Sinica, 17, 63-80

Chen H.S., Simpson D.G. and Ying Z. (2000). Infill asymptotics for a stochastic process model with measurement error. Statistica Sinica 10, 141-156.

Gaetan, C., Guyon, X., 2010. Spatial Statistics and Modeling. Series: Springer Series in Statistics.

Girard, D.A., 2012. Asymptotic near-efficiency of the "Gibbs-energy and empirical-variance" estimating functions for fitting Matérn models to a dense (noisy) series. Preprint version 2 http://arxiv.org/abs/0909.1046

Girard, D.A., 2014. "Estimating a Centered Ornstein-Uhlenbeck Process under Measurement Errors". From Wolfram Demonstrations Project.
http://demonstrations.wolfram.com/EstimatingACenteredOrnsteinUhlenbeckProcessUnderMeasurementE/

Girard, D.A., 2016. Asymptotic near-efficiency of the "Gibbs-energy and empirical-variance" estimating functions for fitting Matérn models - I: Densely sampled processes. Statistics & Probability Letters, 110, 191-197

Girard, D.A., 2017. Efficiently estimating some common geostatistical models by 'energy-variance matching' or its randomized 'conditional-mean' versions. Spatial Statistics, 21(Part A), 1-26.

Girard, D.A., 2018. "Estimators of a Noisy Centered Ornstein-Uhlenbeck Process and Its Noise Variance". From Wolfram Demonstrations Project.
http://demonstrations.wolfram.com/EstimatorsOfANoisyCenteredOrnsteinUhlenbeckProcessAndItsNois/

Heyde, C.C., 1997. Quasi-Likelihood And Its Application: A General Approach to Optimal Parameter Estimation. Springer Series in Statistics

Kaufman, C, Shaby, B., 2013. The role of the range parameter for estimation and prediction in geostatistics. Biometrika, 100, 473-484

Kessler M., 1997. Simple and explicit estimating functions for a discretely observed diffusion process. Scand. J. Statistics 27, 65-82.

Kessler, M., Lindner A., Sorensen, M., 2012. Statistical Methods for Stochastic Differential Equations. Section 1.10 General asymptotic results for estimating functions. CRC Press.

S.C. Kou, On the efficiency of selection criteria in spline regression, Probab. Theory Relat. Fields 127 (2003), pp. 153-176.

Lim, C.Y., Chen C.-H., Wu W.-Y., 2017. Numerical instability of calculating inverse of spatial covariance matrices. Statistics & Probability Letters, 129, 182-188

Stein, M.L., 1999. Interpolation of Spatial Data: Some Theory for Kriging. Springer.

Weisstein, E. W., 2019. "Beta Function." From MathWorld–A Wolfram Web Resource.
http://mathworld.wolfram.com/BetaFunction.html,

Zhang, H., 2004. Inconsistent estimation and asymptotically equivalent interpolation in model-based geostatistics. J. Amer. Statist. Assoc. 99, 250-261.

Zhang H. and Zimmerman D.L., 2005. Toward reconciling two asymptotic frameworks in spatial statistics. Biometrika, 92, 921-936.

Zhang H., Zimmerman D.L., 2007. Hybrid estimation of semivariogram parameters. Mathematical Geology, 39(2), 247-260.

APPENDIX

Let us prove Lemma 2.2 before Lemma 2.1. This can be done by substituting for $a^{\delta}_{b,\theta}$ an "un-aliased" version defined below. So we first establish the following Lemma which states the order of the differences between exact and un-aliased versions:

Lemma A.1. *For any $\lambda \in [-\pi,\pi]$, any $(b,\theta) \in B \times \Theta$, assuming δ bounded, e.g. < 1, there exists constants $\underline{c}_1, \underline{c}_2 > 0$, $\bar{c}_1, \bar{c}_2 < \infty$ only functions of ν, such that:*
1)
$$\underline{c}_1\, \delta^{2\nu} \leq g^{*}_{\nu,\delta\theta}(\lambda) \leq g^{\delta}_{\nu,\theta}(\lambda) \leq g^{*}_{\nu,\delta\theta}(\lambda) + \bar{c}_1\delta^{2\nu},$$
2)
$$\underline{c}_2\, \delta^{2\nu} \leq a^{*}_{b,\delta\theta}(\lambda) \leq a^{\delta}_{b,\theta}(\lambda) \leq a^{*}_{b,\delta\theta}(\lambda) + \bar{c}_2\delta^{2\nu}, \quad \text{where} \quad a^{*}_{b,\cdot} := \frac{b g^{*}_{\nu,\cdot}}{b g^{*}_{\nu,\cdot} + (2\pi)^{-1}}.$$

Proof. Part 2 of Lemma A.1 is a direct consequence of its Part 1 (the second and the third inequalities of Part 2 are immediate consequences after noticing that, for the function $w : x \rightsquigarrow 1/\left(1 + x^{-1}\right)$, we have that $0 < x_1 < x_2$ implies $0 < w(x_2) - w(x_1) < x_2 - x_1$; in the same way, the inequality $w(2\pi b\underline{c}_1\,\delta^{2\nu}) \leq a^*_{b,\delta\theta}(\lambda)$ is a consequence of the first inequality in Part 1; it suffice then to use $\delta < 1$ in $w(2\pi b\underline{c}_1\,\delta^{2\nu}) = 2\pi b\underline{c}_1\,\delta^{2\nu}/\left(2\pi b\underline{c}_1\,\delta^{2\nu}+1\right)$ to obtain the lower bound $\underline{c}_2\delta^{2\nu}$. So let us prove Part 1. The first inequality and the second one are easily checked. Let us prove the third one. Letting $\alpha = \delta\theta$, by arguments identitcal to those used in a step of the proof of Lemma 2.1 in Girard (2016), we can obtain that $\lambda > -\pi$ implies that, for any $k \geq 1$, $g^*_{\nu,\alpha}(\lambda+2\pi k) \leq g^*_{\nu,\alpha}(2\pi(k-1/2)) \leq (C_\nu/(2\pi)^{2\nu+1})\,\alpha^{2\nu}/(k-1/2)^{2\nu+1}$; and summing these terms over $k = 1, 2, \cdots$, thus gives a term $\mathrm{O}(\delta^{2\nu})$ (notice that $\mathrm{O}(\delta^{2\nu+1})$ was obtained in the mentioned step since we considered there $g^*_{\nu,\theta}(\cdot/\delta)$ which can be easily checked to be $\delta g^*_{\nu,\alpha}(\cdot)$). This is shown similarly (except we use $\lambda < \pi$) for the sum over $k = -1, -2, \cdots$. Combining these two results gives the claimed bound for $g^\delta_{\nu,\theta}(\lambda) - g^*_{\nu,\alpha}(\lambda) = \sum_{k\neq 0} g^*_{\nu,\alpha}(\lambda + 2\pi k)$.

Proof of Lemma 2.1 Note that the concise expression $c^k_{2,\nu}$ results from $C_{3/2} = 2\pi {C_{1/2}}^2$. The first equivalence of Lemma 2.1 is trivial by develloping $\int \left[g^\delta_{\nu,\theta}/a^\delta_{b,\theta}\right]^2 = b^{-2}\int\left[bg^\delta_{\nu,\theta} + 1/(2\pi)\right]^2 = \int\left[g^\delta_{\nu,\theta}\right]^2 + 1/(\pi b)\int g^\delta_{\nu,\theta} + 1/(2\pi b)^2$ and recalling that $\int g^\delta_{\nu,\theta} = 1$. As to the second equivalence, let us examine the Proof of Lemma 2.1 of Girard (2016). We make here the same change of variable $\omega = \lambda/\delta$ (note a typographical error: the fraction slash was omitted in Girard (2016)). The integrand is now modified only by the factor $a^\delta_{b,\theta}(\delta\omega)$. Since this factor is < 1, the fact that Lebesgue dominated convergence theorem is applicable has not to be re-proved, and:

$$\begin{aligned}
\delta^{-1}\int_{-\pi}^{\pi}\left(a^\delta_{b,\theta}(\lambda)\frac{g^\delta_{\nu+1,\theta}(\lambda)}{g^\delta_{\nu,\theta}(\lambda)}\right)^2 d\lambda &= \int_{-\pi/\delta}^{\pi/\delta}\left[a^\delta_{b,\theta}(\delta\omega)\right]^2\frac{\left(\sum_{k=-\infty}^{\infty} g^*_{\nu+1,\theta}\left(\omega + 2k\pi/\delta\right)\right)^2}{\left(\sum_{k=-\infty}^{\infty} g^*_{\nu,\theta}\left(\omega + 2k\pi/\delta\right)\right)^2} d\omega \\
&\to \int_{-\infty}^{\infty}\left(\frac{g^*_{\nu+1,\theta}\left(\omega\right)}{g^*_{\nu,\theta}\left(\omega\right)}\right)^2 d\omega,
\end{aligned}$$

the limit being a consequence of $a^\delta_{b,\theta}(\delta\omega) \to 1$ (which can directly be seen from $\delta g^\delta_{\nu,\theta}(\delta\omega) - g^*_{\nu,\theta}\left(\omega\right) \to 0$ as $\delta \to 0$, this difference being shown in fact $O(\delta^{2\nu})$ in the Proof of Lemma 2.1 of Girard (2016)).

Proof of Lemma 2.2. By Lemma A.1, since $\left|\left[a^\delta_{b,\theta}(\lambda)\right]^k - \left[a^*_{b,\delta\theta}(\lambda)\right]^k\right| < k\left|a^\delta_{b,\theta}(\lambda) - a^*_{b,\delta\theta}(\lambda)\right| = O(\delta^{2\nu})$ one can replace, with an accuracy which will be sufficient since $2\nu > \frac{2\nu}{2\nu+1}$, the filter by its un-aliased version in $\int\left[a^\delta_{b,\theta}(\lambda)\right]^k \mathrm{d}\lambda$. On the other hand, by the change of variable $s = \lambda/(\delta\theta)^{\frac{2\nu}{2\nu+1}}$ and an application of the dominated convergence theorem, one can find that

$$(\delta\theta)^{-\frac{2\nu}{2\nu+1}}\int_0^{2\pi}\left[a^*_{b,\delta\theta}(\lambda)\right]^k \mathrm{d}\lambda \to \int_0^\infty (1 + (2\pi C_\nu b)^{-1}s^{2\nu+1})^{-k}\mathrm{d}s = (2\pi C_\nu b)^{\frac{1}{2\nu+1}}\int_0^\infty (1 + s^{2\nu+1})^{-k}\mathrm{d}s.$$

The claimed equivalents are now obtained from a known expression, for $k \in \{1, 2\}$, of the latter classic integral.

Proof of Theorem 3.3. It is convenient to apply Theorem 1.58 of Kessler et al. (2012) which is a general asymptotic-consistency result for estimating equation. In fact, our Theorem 3.1 exactly states that the conditions required to apply their Theorem 1.58 are well fulfilled for, in their notations (except we set here $\overline{\theta} := (b, \theta)^T$), the two-component estimating equation $G_n\left(\overline{\theta}\right) :=$

$(\sigma_{\rm N}^{-2}\mathbf{y}^T\mathbf{y}/n - b - 1, \sigma_{\rm N}^{-2}\mathrm{CGEM}(b,\theta) - b)^T$. Precisely, in their notations, the required conditions (i) and (ii) (resp. (iii)) are immediate consequence of Part 1 (resp. Part 3) of Theorem 3.1.

Proof of Part 1 of Theorem 4.1. We shall apply Theorem 1.60 of Kessler et al. (2012) to the two-component equation $G_n\left(\overline{\theta}\right)$ defined above. It is clear, form the high differientiabily regularity, already mentioned, of $g^{\delta}_{\nu,\theta}$ and its strict positivity, that G_n is continuously differentiable over $B\times\Theta$. Denoting by $\frac{\partial}{\partial\overline{\theta}^T}\overline{G}_n(b,\theta)$ the Jacobian matrix, letting $M_{\delta,b_0,\theta_0}(b,\theta) :=$

$$\left[\begin{array}{cc} -1 & 0 \\ \frac{-1}{2\pi}\left(\int\left\{[a^{\delta}_{b,\theta}]^2 - 2\left([a^{\delta}_{b,\theta}]^2 - [a^{\delta}_{b,\theta}]^3\right)\left(\frac{b_0g_0}{bg^{\delta}_{\nu,\theta}} - 1\right)\right\}\right) & \frac{-b}{2\pi}\left(\int\left\{[a^{\delta}_{b,\theta}]^2 + \left(2[a^{\delta}_{b,\theta}]^3 - [a^{\delta}_{b,\theta}]^2\right)\left(\frac{b_0g_0}{bg^{\delta}_{\nu,\theta}} - 1\right)\right\}h^{\delta}_{\nu,\theta}\right)\end{array}\right],$$

it is a consequence of Theorem 3.1 that $\frac{\partial}{\partial\overline{\theta}^T}\overline{G}_n(b,\theta) \to M_{\delta,b_0,\theta_0}(b,\theta)$, uniformly over $B\times\Theta$ and $M_{\delta,b_0,\theta_0}(b_0,\theta_0)$ is invertible. Furthermore, $n^{1/2}\overline{G}_n\left(\overline{\theta}_0\right) \xrightarrow{\mathcal{D}} N\left(\left[\begin{array}{c}0\\0\end{array}\right], 2b_0^2\left[\begin{array}{cc}2\pi\int a_0^{-2}g_0^2 & 1\\ 1 & \frac{1}{2\pi}\int a_0^2\end{array}\right]\right)$ can be deduced from classic time-series results (e.g. in Azencott and Dacunha-Castelle (1986)) since the required regularity conditions are well fulfilled (as discussed in Section 3 and in Girard (2016))). Then Theorem 1.60 of Kessler et al. (2012) is applicable and it gives (after direct, albeit tedious, algebraic manipulations) that the asymptotic covariance matrix of $(\hat{b}_{\mathrm{EV}|\sigma_{\rm N}}, \hat{\theta}_{\mathrm{GEV}|\sigma_{\rm N}})$ is

$$4\pi\left[\begin{array}{cc} b_0^2\int a_0^{-2}g_0^2 & b_0\frac{1-\left(\int a_0^2\int a_0^{-2}g_0^2\right)}{\int a_0^2h_0} \\ b_0\frac{1-\left(\int a_0^2\int a_0^{-2}g_0^2\right)}{\int a_0^2h_0} & \int a_0^2\frac{\left(\int a_0^2\int a_0^{-2}g_0^2\right)-1}{\left|\int a_0^2h_0\right|^2}\end{array}\right].$$