Statistics and Probability Letters 130 (2017) 63–70

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Parametric inference of autoregressive heteroscedastic models with errors in variables

Salima El Kolei [a,*], Florian Pelgrin [b]

[a] CREST, ENSAI, UBL Campus de Ker-Lann, rue Blaise Pascal, BP 37203, 35172 Bruz cedex, France
[b] EDHEC Business School, 24, avenue Gustave Delory CS 50411, 59057 Roubaix Cedex 1, France

ARTICLE INFO

Article history:
Received 29 January 2017
Received in revised form 3 July 2017
Accepted 21 July 2017
Available online 31 July 2017

Keywords:
Hidden Markov model
Parametric estimation
Deconvolution
Heteroscedasticity

ABSTRACT

We propose a consistent and asymptotically normal parametric estimator for autoregressive heteroscedastic models with errors in variables based on contrast minimization and give an example for a discrete time observed CIR process with additive noises.

1. Introduction

This paper is motivated by the parametric estimation of hidden stochastic models of the form:

$$\begin{cases} Y_i = X_i + \varepsilon_i \\ X_{i+1} = b_{\theta_0}(X_i) + \sigma_{\theta_0}(X_i)\eta_{i+1}, \end{cases} \quad (1)$$

where one observes $Y_1, \ldots, Y_n$, and where the random variables ε_i, η_i and X_i are unobserved. Notably $(X_i)_{i\geq 0}$ is a strictly stationary, ergodic process that depends on two measurable functions b_{θ_0} and σ_{θ_0} and its stationary density is f_{θ_0}, where θ_0 belongs to $\Theta \subset \mathbb{R}^p$. The functions b_{θ_0}, σ_{θ_0} and f_{θ_0} are known up to a finite dimensional parameter, θ_0. Finally, the innovations $(\eta_i)_{i\geq 0}$ and the errors $(\varepsilon_i)_{i\geq 0}$ are independent and identically distributed (i.i.d.) random variables, the distribution of the noises $(\varepsilon_i)_{i\geq 0}$ being known for identifiability of the model.

In this work, we propose to estimate the parameters of the two functions b_{θ_0} and σ_{θ_0} driving the dynamics of the hidden variables $(X_i)_{i\geq 0}$. Our method extends the previous work of El Kolei (2013) where a contrast approach is proposed for homoscedastic noises.

In many applications of interest, the assumption of homoscedastic errors is too restrictive to be realistic. In Delaigle and Meister (2008), the authors introduce a kernel estimator of the density f_{θ_0} in the case of heteroscedastic contamination.

In this paper, the heteroscedasticity appears in the unobserved component which is also a property observed in practice (financial, biology, chemistry). For this purpose, we define a new contrast function and under mild assumptions we show that our estimator is consistent and asymptotically normally distributed which leads to obtain Confidence Intervals (CI) in practice. It is worth noticing that our approach relies on Comte et al. (2010) where the authors propose a nonparametric Nadaraya–Watson estimator of the two functions b_{θ_0} and σ_{θ_0}. In the same perspective, in Dedecker et al. (2014) the authors propose a semi-parametric estimator of θ_0 based on a weighted least square estimation for homoscedastic noises. Their

* Corresponding author.
E-mail address: salima.el-kolei@ensai.fr (S. El Kolei).

http://dx.doi.org/10.1016/j.spl.2017.07.011
0167-7152/

estimator is based on the introduction of a kernel deconvolution density and depends on the choice of a weight function. The approach proposed here is different: it is not based on a weighted least square estimation and allows to estimate nonlinear autoregressive models with heteroscedastic noises.

The fields of application are various and include astronomy, biology, chemistry, economy; see the numerous examples described in Carroll et al. (1995). For practical issue, we propose to illustrate our estimator on a CIR process (Cox Ingersoll Ross, see Cox et al., 1985) since it has found many applications and since this process presents serious heteroscedasticity. We provide an analytical expression of the contrast function for this process, and we compare our estimator with a Monte Carlo Expectation Maximization Likelihood estimator (MCEML) for different number of observations and various types of errors distribution. Our Monte Carlo simulations show that our approach gives results similar to the MCEML but our approach is the fastest computing. Furthermore, our estimator has the good expected properties (unbiased and small MSE) and presents good convergence properties. The errors distribution seems to have a slight influence on the MSE which is related to the current theoretical properties of deconvolution (see Fan, 1991).

The paper is organized as follows. Section 2 presents the notations and the model assumptions. Section 3 defines the deconvolution-based M-estimator and states all of the theoretical properties. Some Monte Carlo simulations are discussed in Section 4.

2. General setting and assumptions

2.1. Notations

Subsequently, for any function $v : \mathbb{R} \to \mathbb{R}$, we denote by v^* the Fourier transform of the function v: $v^*(t) = \int e^{itx} v(x)dx$, by $\|v\|$ its $\mathbb{L}_2(\mathbb{R})$-norm, $\langle \cdot, \cdot \rangle$ stands for the scalar product in $\mathbb{L}_2(\mathbb{R})$ and "$\star$" for the usual convolution product. Moreover, for any integrable and square-integrable functions u, u_1, and u_2: we have $(u^*)^*(x) = 2\pi u(-x)$ and $\langle u_1, u_2 \rangle = \frac{1}{2\pi} \langle u_1^*, u_2^* \rangle$. Finally, $\|A\|$ denotes the Euclidean norm of a matrix A, $\mathbf{Y}_i = (Y_i, Y_{i+1})$ and $\mathbf{y}_i = (y_i, y_{i+1})$, $\mathbf{P}_n$ (respectively, $\mathbf{P}$) the empirical (respectively, theoretical) expectation, that is, for any stochastic variable: $\mathbf{P}_n(X) = \frac{1}{n}\sum_{i=1}^n X_i$ (respectively, $\mathbf{P}(X) = \mathbb{E}[X]$). Regarding the partial derivatives, for any function h_θ, $\nabla_\theta h_\theta$ is the vector of the partial derivatives of h_θ with respect to (w.r.t) θ and $\nabla_\theta^2 h_\theta$ is the Hessian matrix of h_θ w.r.t θ.

2.2. Assumptions

A0: θ_0 belongs to the interior Θ_0 of a compact set Θ, $\theta_0 \in \Theta \subset \mathbb{R}^p$; **A1**: the errors $(\varepsilon_i)_{i\geq 0}$ are independent and identically distributed centered random variables with finite variance, $\mathbb{E}\left[\varepsilon_1^2\right] = s_\varepsilon^2$. The density of ε_1, f_ε, belongs to $\mathbb{L}_2(\mathbb{R})$, and for all $x \in \mathbb{R}$, $f_\varepsilon^*(x) \neq 0$; **A2**: the innovations $(\eta_i)_{i\geq 0}$ are independent and identically distributed centered random variables; **A3**: the X_i's are strictly stationary, ergodic and α-mixing with invariant density f_{θ_0}; **A4**: the sequences $(X_i)_{i\geq 0}$ and $(\varepsilon_i)_{i\geq 0}$ are independent. The sequence $(\varepsilon_i)_{i\geq 0}$ and $(\eta_i)_{i\geq 0}$ are independent; **A5**: on Θ_0, the functions $\theta \mapsto b_\theta$ and $\theta \mapsto \sigma_\theta$ admit continuous derivatives with respect to θ up to order 2; **A6**: The function to estimate $l_\theta := \left(b_\theta^2 + \sigma_\theta^2\right) f_\theta$ belongs to $\mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R})$, is twice continuously differentiable w.r.t $\theta \in \Theta$ for any x and measurable w.r.t x for all θ in Θ. Each element of $\nabla_\theta l_\theta$ and $\nabla_\theta^2 l_\theta$ belongs to $\mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R})$; **A7**: the application $\theta \mapsto \mathbf{P}m_\theta$ admits a unique minimum and its Hessian matrix, denoted by V_θ, is non-singular in θ_0.

Remark 1. The compactness assumption **A0** might be relaxed by assuming that θ_0 is an element of the interior of a convex parameter space $\Theta \in \mathbb{R}^p$. Assumptions **A1**–**A2** are quite standard when considering estimation for convolution models. Assumption **A3** is useful for the statistical properties of our estimator. We give just below some conditions on the functions b, σ and η ensuring this assumption. Assumptions **A5**–**A6** ensure some smoothness for the functions b, σ. Assumption **A7** is also quite usual in the literature and serves for the construction and for asymptotic properties of our estimator.

Let us consider the process $(X_i)_{i\geq 0}$ defined in (1), we give conditions on the functions b, σ and η ensuring that assumption **A3** is satisfied.

(**i**) The random variables $(\eta_i)_{i\geq 0}$ are i.i.d. with an everywhere positive and continuous density function independent of (X_i).
(**ii**) The function b_{θ_0} is bounded on every bounded set; that is, for every $K > 0$, $\sup_{|x|\leq K} |b_{\theta_0}(x)| < \infty$.
(**iii**) The function σ_{θ_0} satisfies, for every $K > 0$ and constant σ_1, $0 < \sigma_1 \leq \inf_{|x|\leq K} \sigma_{\theta_0}(x)$ and $\sup_{|x|\leq K} \sigma_{\theta_0}(x) < \infty$.
(**iv**) There exist constants $C_b > 0$ and $C_\sigma > 0$, sufficiently large $M_1 > 0$, $M_2 > 0$, $c_1 \geq 0$ and $c_2 \geq 0$ such that $|b_{\theta_0}(x)| \leq C_b|x| + c_1$, for $|x| \geq M_1$ and $|\sigma_{\theta_0}(x)| \leq C_\sigma|x| + c_2$, for $|x| \geq M_2$ and $C_b + \mathbb{E}[\eta_1]C_\sigma < 1$.

Under assumptions (**i**)–(**iv**), the process $(X_i)_{i\geq 0}$ defined in (1) is strictly stationary and strongly mixing with a geometric convergence rate (see Doukhan, 1994 and Chen and Chen, 2000).