Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 137 (2007) 1627 – 1633
journal of statistical planning and inference
www.elsevier.com/locate/jspi

An alternative derivation of the Kalman filter using the quasi-likelihood method

Yan-Xia Lin*

School of Mathematics and Applied Statistics, University of Wollongong, Australia

Available online 9 October 2006

Abstract

The Kalman filter gives a recursive procedure for estimating state vectors. The recursive procedure is determined by a matrix, so-called gain matrix, where the gain matrix is varied based on the system to which the Kalman filter is applied. Traditionally the gain matrix is derived through the maximum likelihood approach when the probability structure of underlying system is known. As an alternative approach, the quasi-likelihood method is considered in this paper. This method is used to derive the gain matrix without the full knowledge of the probability structure of the underlying system. Two models are considered in this paper, the simple state space model and the model with correlated between measurement and transition equation disturbances. The purposes of this paper are (i) to show a simple way to derive the gain matrix; (ii) to give an alternative approach for obtaining optimal estimation of state vector when underlying system is relatively complex.

Keywords: The Kalman filter; Quasi-likelihood; Quasi-score estimating function; Quasi-likelihood estimate

1. Introduction

The Kalman filter is an important algorithm and has many applications in information processes, economics and repeated sample surveys (see Feder, 2001; Maybeck, 1979; Harvey, 1994 and reference therein). It is a recursive procedure. A matrix, so-called gain matrix, is involved in the procedure and plays an important role in the procedure. When the probability structure of the system of interest is known, especially the structure is normal or conditional normal, the gain matrix is usually derived by using the maximum likelihood method (see, Harvey, 1994). Some literature also discuss how to derive the formula of gain matrix for other situations, for examples, a linear system without full knowledge on probability structure or a nonlinear system with normal or conditional normal probability structure (see Anderson and Moore, 1979; Duncan and Horn, 1972). Basically, these driving procedures are tedious and complex.

The quasi-likelihood (QL) technique has a lot of applications on the inference of stochastic processes and it is used to obtain optimal estimation of parameters in the stochastic process of interest (see, Heyde, 1997). From inference point of view, the Kalman filter procedure is also an estimation procedure. It claims that the output given by the Kalman filter is optimal. Both of procedures are related to optimum issues. There must be some relationship between them. However, due to lack of communication, this relationship has not been well studied.

* Tel.: +61 42 214336; fax: +61 42 214845.
E-mail address: yanxia@uow.edu.au.

0378-3758/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2006.09.010

In this paper, the QL method is used to derive the gain matrix for two models: the simple state space model and the model with correlated between measurement and transition equation disturbances. It shows that the Kalman filter for these two models is a special application of the QL method to the models. It also shows that, as an alternative approach, the procedure of using the QL approach to obtain the gain matrix is fairly standard and easily to be followed.

2. The QL method

Let $\{y_t, 0 \leqslant t \leqslant T\}$ be a sample in discrete or continuous time drawn from some process taking values in r-dimensional Euclidean space whose distribution depends on a parameter θ taking value in an open subset Θ of p-dimensional Euclidean space. Suppose that the possible probability distribution functions for $\{y_t\}$ are $\mathscr{P}=\{\mathscr{P}_\theta\}$, a union of parametric families, each family being indexed by the same parameter θ and that each $(\Omega, \mathscr{F}, P_\theta)$ is a complete probability space, where Ω is the population for $\{y_t\}$ and $\mathscr{F}$ is a σ-field generated by $\{y_t\}$. Denote $\mathscr{F}_t$ a σ-field generated by $\{y_l\}_{l\leqslant t}$, that is, $\mathscr{F}_t$ contains all the information provided by the process $\{y_l\}$ up to time t.

Let Ψ_T be a class of mean zero, square integrable functions $G_T = G_T(\{y_t, 0\leqslant t\leqslant T\}, \theta)$, that is $EG_T(\theta) = 0$ and $E[G_T G_T'] < \infty$, for each $P_\theta \in \mathscr{P}$ with index θ, where $'$ denotes transpose. Let $\mathscr{G}_T$ be a subset of Ψ_T in which every element G_T is almost surely differentiable with respect to the components of θ and for which $E[\dot{G}_T(\theta)] = (E[\partial G_{T,i}(\theta)/\partial\theta_j])_{p\times r}$ and $E[G_t(\theta)G_t'(\theta)]$ are nonsingular. The expectations are always with respect to P_θ.

Following the above notation, Godambe and Heyde (1987) gave the following definition for a quasi-score estimating function in space $\mathscr{G}_T$.

Definition 1. $G_T^*(\theta)$ is a quasi-score estimating function in $\mathscr{G}_T \subseteq \Psi_T$, if

$$(E\dot{G}_T)^{-1}E(G_T G_T')(E\dot{G}_T')^{-1} - (E\dot{G}_T^*)^{-1}E(G_T^* G_T^{*'})(E\dot{G}_T^{*'})^{-1}$$

is nonnegative-definite for all $G_T(\theta) \in \mathscr{G}_T$, $\theta \in \Theta$ and $P_\theta \in \mathscr{P}$.

Equation $G_T^*(\theta) = 0$ is called a quasi-score normal equation. If $G_T^*(\theta)$ satisfies that $(E\dot{G}_T^*)^{-1}E(G_T^* G_T^{*'}) = 1$, we call G_T^* a standardized quasi-score estimating function (associated with the space $\mathscr{G}_T$). The solution of $G_T^*(\theta) = 0$ is called the QL estimate of θ.

The QL method was first introduced by Wedderburn (1974). Wedderburn's work was mainly based on the generalized linear model. At the same time, a similar technique was also independently developed by Godambe and Heyde. This technique latter was called "QL" (see, Godambe and Heyde, 1987). The later technique is more focused on the applications to the inference of stochastic processes. These two independently developed QL methods are defined in different ways because the original approaches are different. The definition given by Godambe and Heyde (1987) is more general than that given by Wedderburn (1974). For this aspect of discussion see Lin and Heyde (1993). We adopt the definition of the QL given by Godambe and Heyde (1987) in this paper.

Consider a stochastic process $y_t \in R^r$,

$$y_t = \mu_t(\theta) + m_t, \quad 0\leqslant t\leqslant T, \tag{1}$$

where $\theta \in \Theta \in R^p$ is the parameter needed to be estimated; μ_t is a function of $\{y_s\}_{s<t}$; (in the other words, μ_t is $\mathscr{F}_{t-1}$-measurable); m_t is an error process with $E(m_t|\mathscr{F}_{t-1}) = E_{t-1}(m_t) = 0$. When the following estimating function space

$$\mathscr{G}_T = \left\{ \sum_{t=1}^{T} \alpha_t(y_t - \mu_t) \mid \alpha_t \text{ is a } \mathscr{F}_{t-1}\text{-measurable } r \times p \text{ matrix} \right\}$$

is considered, the standardized quasi-score estimating function in the space has the following form:

$$G_T^*(\theta) = \sum_{i=1}^{T} (E_{t-1}(\dot{m}_t))'(E_{t-1}(m_t m_t'))^{-1} m_t.$$