Journal de la Société Française de Statistique
Vol. 153 No. 1 (2012)

Asymptotically efficient statistical predictors

Titre : Prédicteurs statistiques asymptotiquement efficaces

Denis Bosq[1] and Emmanuel Onzon[2]

Abstract: Asymptotic efficiency is an optimality criteria for estimators in the theory of point estimation. Here we present an extension of this criteria to plug-in predictors, under the quadratic risk, for parametric prediction problems.

First, asymptotic efficiency is defined for the quadratic error of estimation of the regression function (QER). Under suitable conditions, the QER is asymptotically equivalent to the quadratic error of prediction (QEP). A definition of asymptotic efficiency for predictors is deduced.

Then, a result about limit in distribution for predictors is proved. This yields an alternative definition of asymptotic efficiency for predictors. The results are applied to the problem of forecasting the Ornstein-Uhlenbeck process throughout the paper and simulation results are presented.

Résumé : L'efficacité asymptotique est un critère d'optimalité pour les estimateurs dans la théorie de l'estimation ponctuelle. Nous présentons ici une extension de ce critère pour les prédicteurs plug-in, sous risque quadratique, pour des problèmes de prédiction paramétriques.

On commence par définir l'efficacité asymptotique pour l'erreur quadratique d'estimation de la fonction de régression (EQR). Sous certaines conditions, l'EQR est asymptotiquement équivalente à l'erreur quadratique de prédiction (EQP). On en déduit une définition de l'efficacité asymptotique pour les prédicteurs.

Un résultat de limite en loi pour les prédicteurs est démontré. Une définition alternative de l'efficacité asymptotique pour les prédicteurs en découle. Les résultats sont appliqués au problème de prévision du processus de Ornstein-Uhlenbeck tout au long de l'article et des résultats de simulation sont présentés.

Keywords: forecasting, prediction, regression, asymptotic efficiency, Ornstein-Uhlenbeck process
Mots-clés : prévision, prédiction, régression, efficacité asymptotique, processus d'Ornstein-Uhlenbeck
AMS 2000 subject classifications: 62M20, 62F12, 62J02

1. Introduction

In this paper we are interested in predicting an unknown real random variable Y based on a known random variable X, assuming the couple (X,Y) follows the distribution P_θ where θ is an unknown parameter that belongs to the domain $\Theta \subset \mathbb{R}$. If $p(X)$ is a predictor of Y, then we measure its performance by the quadratic error of prediction (QEP hereafter) which breaks down into the sum of the following two terms

$$\mathrm{E}_\theta \left(p(X)-Y\right)^2 = \mathrm{E}_\theta \left(p(X)-\mathrm{E}_\theta[Y|X]\right)^2 + \mathrm{E}_\theta \left(\mathrm{E}_\theta[Y|X]-Y\right)^2.$$

We call $\mathrm{E}_\theta \left(p(X)-\mathrm{E}_\theta[Y|X]\right)^2$ the *statistical* prediction error and $\mathrm{E}_\theta \left(\mathrm{E}_\theta[Y|X]-Y\right)^2$ the *probabilistic* prediction error. We see that minimizing the QEP of $p(X)$ as a predictor of Y is the

[1] LSTA, Université Pierre et Marie Curie.
E-mail: denis.bosq@upmc.fr
[2] LSTA, Université Pierre et Marie Curie.
E-mail: emmanuel.onzon@upmc.fr

same as minimizing the statistical prediction error, which is the QEP of $p(X)$ as a predictor of $r(X,\theta) = \mathrm{E}_\theta[Y|X]$. Thus, for the prediction problem under quadratic loss, the quantity of interest in its most general form is $r(X,\theta)$. And in what follows, the QEP will always refer to the error with respect to $r(X,\theta)$.

It is natural to see the function $x \mapsto r(x,\theta)$ as the regression function with regressor X and regressand Y. The problem of estimation of the regression function $r(\cdot,\theta)$ is related to the problem of prediction of $r(X,\theta)$ since if $\widehat{r}$ is an estimator of the function $r(\cdot,\theta)$ then $p(X) = \widehat{r}(X)$ can be seen as a predictor of $r(X,\theta)$. Nonetheless they are two distinct problems since the regression function $r(\cdot,\theta)$ is an unknown *deterministic* quantity while $r(X,\theta)$ is an unknown *random* element. This is why the risk of a predictor and the risk of an estimator of the regression function are measured differently. We define the quadratic error with respect to the regression function (QER hereafter) of $\widehat{r}$ to estimate $r(\cdot,\theta)$ as follows

$$\rho(\theta) = \int_{\mathscr{X}} \mathrm{E}_\theta \left(\widehat{r}(x) - r(x,\theta)\right)^2 d\mu_\theta(x),$$

where we choose μ_θ, a measure on $(\mathscr{X},\mathscr{A})$, with $\mathscr{X}$ the range of the random variable X. Oftentimes μ_θ will be the distribution of X under P_θ. Here $\widehat{r}$ is a random function based on X such that for all $x \in \mathscr{X}$, $\widehat{r}(x)$ is an estimator of $r(x,\theta)$ (in some sense the dependence of $\widehat{r}(X)$ on X is twofold). If the function r is known then $\widehat{r}(x)$ can take the form $\widehat{r}(x) = r(x,\widehat{\theta})$ where $\widehat{\theta}$ is an estimator of θ, we will call such estimators plug-in estimators.

One approach to the problem of prediction is the extension of the theory of UMVUE (uniformly minimum variance unbiased estimation, see for instance [8]) to the case of statistical prediction (see [14], [9], [3] ch. 1). In particular a Cramér-Rao type lower bound has been derived for the QEP (see [14], [10], [11]), and predictors which QEP attain the bound are said to be *efficient* predictors in an analogy with efficient estimators. This information inequality reads

$$\mathrm{E}_\theta(p(X) - r(X,\theta))^2 \geqslant b(\theta)^2 + \frac{(b'(\theta) + \mathrm{E}_\theta[\partial_\theta r(X,\theta)])^2}{I(\theta)},$$

where $b(\theta) = \mathrm{E}_\theta p(X) - \mathrm{E}_\theta r(X,\theta)$ is the bias of $p(X)$ to predict $r(X,\theta)$ and $I(\theta)$ the Fisher information of X. Unsurprisingly this lower bound for predictors suffers the same drawbacks as the Cramér-Rao bound for estimators. More often than not it is not attainable and it depends on the bias of the predictor. In the case of point estimation, a theory of asymptotic efficiency for sequences of estimators has been developped, in order to go past these limitations (see for instance [5], [7], [12] and [8] ch. 6). Roughly speaking, this theory gives asymptotic lower bounds on the (normalized) risk of the estimator when the number of observations tends to the infinity.

In this article we attempt to follow such an approach for sequences of predictors and estimators of the regression function. We are interested in dependent data such as time series indexed by $\mathbb{N}$ or random processes indexed by $\mathbb{R}_+$. We will denote by $X_{(T)} = (X_t, 0 \leqslant t \leqslant T)$ the observations until time $T > 0$ with X_t taking its values in a measurable space $(E,\mathscr{B})$. We will consider problems of estimation of the function $r_T(\cdot,\theta)$ and of prediction of $r_T(Z_T,\theta)$ given $X_{(T)}$, where Z_T is some $X_{(T)}$-measurable random variable taking values in some measurable space $(\mathscr{Z}_T,\mathscr{C}_T)$, and $r_T(\cdot,\theta)$ is some function defined on $\mathscr{Z}_T$ and taking values in $\mathbb{R}$. For instance, assuming $E = \mathbb{R}$, one may want to predict $r_T(Z_T,\theta) = \mathrm{E}_\theta[X_{T+h}|X_{(T)}]$ for some $h > 0$. We will restrict our attention to plug-in predictors, *i.e.* predictors of the form $p(Z_T) = r_T(Z_T,\widehat{\theta}_T)$ where $\widehat{\theta}_T$ is an estimator of the parameter θ based on $X_{(T)}$. In sections 2 and 3, we assume the process $(X_t, t \geqslant 0)$ is Markov, hence $Z_T = X_T$. In this setting, the QEP and the QER will refer to the following respective quantities.

$$R_T(\theta) = \mathrm{E}_\theta \left(r_T(X_T, \widehat{\theta}_T) - r_T(X_T, \theta) \right)^2,$$

and

$$\rho_T(\theta) = \int_E \mathrm{E}_\theta \left(\widehat{r}_T(x) - r_T(x, \theta) \right)^2 d\mu_{\theta,T}(x),$$

where $\widehat{r}_T$ is an estimator of the regression function $r_T(\cdot, \theta)$, that may be the plug-in estimator $\widehat{r}_T = r_T(\cdot, \widehat{\theta}_T)$ and $\mu_{\theta,T}$ is a measure on $(E, \mathscr{B})$, for instance the distribution of X_T.

To motivate our approach, let us illustrate the limitations of the information inequality on the following prediction problem. Let $(X_t, t \geqslant 0)$ be a stationary Ornstein-Uhlenbeck process with unknown drift coefficient $\theta \in \mathbb{R}_+^*$ and known diffusion coefficient $\sigma = 1$. We consider the prediction of X_{T+h} given $X_{(T)}$. The conditional expectation is $\mathrm{E}_\theta[X_{T+h} | X_{(T)}] = e^{-\theta h} X_T = r_T(X_T, \theta)$. The expectation of its derivative vanishes, since the process is centered, $\mathrm{E}_\theta[\partial_\theta r_T(X_T, \theta)] = \mathrm{E}_\theta[-h e^{-\theta h} X_T] = 0$. Therefore the bound for an unbiased predictor is zero, *i.e.* an unbiased efficient predictor is equal to the conditional expectation $e^{-\theta h} X_T$ which is not possible since it depends on the unknown parameter θ. Hence the bound is not attainable. Even worse, no uniformly best unbiased predictor exists since for any fixed $\theta_0 \in \mathbb{R}_+^*$, $p_0(X_T) = e^{-\theta_0 h} X_T$ is an optimal unbiased predictor at θ_0. Here it will be more fruitful to consider the behavior of the predictor when $T \to \infty$. We will see that for this problem, plug-in predictors based on asymptotically efficient estimators of θ, are asymptotically efficient predictors.

We will consider two different ways of defining asymptotic efficiency for plug-in predictors in section 3 and section 4 respectively.

In section 2 we study asymptotic efficiency for the estimation of the regression function. In section 3, the first definition of asymptotic efficiency for plug-in predictors follows from asymptotic equivalence of the QER and the QEP, under suitable conditions. While the second definition follows, in section 4, from a result about the limit in distribution of the predictor.

In section 2 we consider the problem of estimation of the regression function $r_T(\cdot, \theta)$. We assume the process $(X_t, t \geqslant 0)$ is Markov, hence the function $r_T(\cdot, \theta)$ is defined on the space E, where the process $(X_t, t \geqslant 0)$ takes its values. Then under suitable assumptions on the process and on the estimator $\widehat{r}_T$, the limit of the QER is bounded from below as follows,

$$\varliminf_{T \to \infty} \frac{\rho_T(\theta)}{v_T(\theta)} \geqslant 1, \quad \text{with} \quad v_T(\theta) = \frac{\|\partial_\theta r_T(\cdot, \theta)\|^2_{\mu_{\theta,T}}}{I_T(\theta)},$$

where $I_T(\theta)$ is the Fisher information of $X_{(T)}$, and $\|\cdot\|_{\mu_{\theta,T}}$ denotes the norm of $\mathrm{L}^2(\mu_{\theta,T})$. Moreover for plug-in estimators $r_T(\cdot, \widehat{\theta}_T)$, and under additional assumptions,

$$\lim_{T \to \infty} T \rho_T(\theta) = U(\theta) V(\theta),$$

where

$$U(\theta) = \lim_{T \to \infty} \|\partial_\theta r_T(\cdot, \theta)\|^2_{\mu_{\theta,T}} \quad \text{and} \quad V(\theta) = \lim_{T \to \infty} T \mathrm{E}_\theta(\widehat{\theta}_T - \theta)^2.$$

Now in section 3, asymptotic equivalence of the QER and the QEP is derived, under some conditions, when $\mu_{\theta,T}$ is the distribution of X_T. This will allow to define asymptotic efficiency for a plug-in predictor, under the assumptions of the results, as

$$T\mathrm{E}_\theta(\widehat{\theta}_T-\theta)^2 \xrightarrow[T\to\infty]{} I(\theta)^{-1},$$

where $I(\theta)=\lim_{T\to\infty}\frac{I_T(\theta)}{T}$.

In a second attempt to define asymptotic efficiency for plug-in predictors we consider, in section 4, the limit in distribution of $\sqrt{T}\big(r_T(Z_T,\widehat{\theta}_T)-r_T(Z_T,\theta)\big)$. Here we do not assume the process is Markov but that Z_T is $\sigma\big(X_t,\phi(T)\leqslant t\leqslant T\big)$-measurable, with $\frac{\phi(T)}{T}\to 1$. Under suitable assumptions,

$$\sqrt{T}\big(r_T(Z_T,\widehat{\theta}_T)-r_T(Z_T,\theta)\big) \xrightarrow[T\to\infty]{d} UV,$$

where $\partial_\theta r_T(Z_T,\theta) \xrightarrow[T\to\infty]{d} U$ and $\sqrt{T}(\widehat{\theta}_T-\theta) \xrightarrow[T\to\infty]{d} V$, with U and V independent random variables. When the assumptions are fulfilled, this result allows to define asymptotic efficiency of the plug-in predictor as being equivalent to the asymptotic efficiency of the estimator $\widehat{\theta}_T$.

2. Asymptotic efficiency for the QER

2.1. Model

Let $(X_t,t\geqslant 0)$, where $t\in\mathbb{N}$ or $t\in\mathbb{R}_+$, be a Markov process taking its values in E. Its distribution is denoted by P_θ where $\theta\in\Theta$ is an unknown real parameter.

For any $T>0$, we consider the problem of estimation of a regression function $E\to\mathbb{R}$, $x\mapsto r_T(x,\theta)$ given the past values of the process until time T,

$$X_{(T)}=(X_t,0\leqslant t\leqslant T).$$

For instance $r_T(\cdot,\theta)$ may be the following conditional expectation.

$$x\mapsto r_T(x,\theta)=\mathrm{E}_\theta[Y|X_T=x],$$

with Y a real random variable which is $\sigma(X_t,t\geqslant T+h)$ measurable and $h>0$.

We consider the following QER.

$$\rho_T(\theta)=\int_E \mathrm{E}_\theta\big(\widehat{r}_T(x)-r_T(x,\theta)\big)^2 d\mu_{\theta,T}(x),$$

where we choose $\mu_{\theta,T}$, a measure on $(E,\mathscr{B})$, with E the range of the random variable X_T, and $\widehat{r}_T$ is an estimator of the regression function $r_T(\cdot,\theta)$. Oftentimes $\mu_{\theta,T}$ will be the distribution of X_T under P_θ, and $\widehat{r}_T$ will be a plug-in estimator $\widehat{r}_T=r(\cdot,\widehat{\theta}_T)$. We will always assume that the function $(x,\theta)\mapsto r_T(x,\theta)$ is measurable for any $T\geqslant 0$.

2.2. Information inequality for the QER

We denote by $\mathscr{X}_T$ the space of the sample paths of $X_{(T)}$. We make the following assumptions.

Assumption 2.1. *For any $T \geqslant 0$ and any $\theta \in \Theta$, the following conditions hold.*

1. *$\Theta \subset \mathbb{R}$ is an open set.*
2. *The distribution of $X_{(T)}$ under P_θ has density $f_T(\cdot,\theta)$ with respect to some σ-finite measure ν.*
3. *The set $\{\xi \in \mathscr{X}_T | f_T(\xi,\theta) > 0\}$ does not depend on θ.*
4. *The derivative $\partial_\theta f_T(\xi,\theta)$ exists and is finite for any $\xi \in \mathscr{X}_T$.*
5. *The equality $\int_{\mathscr{X}_T} f_T(X_{(T)},\theta)d\nu(\xi) = 1$ is differentiable under the $\int$ sign.*
6. *$I_T(\theta) = \mathrm{E}_\theta\left(\partial_\theta \log f_T(X_{(T)},\theta)\right)^2 \in (0,\infty)$.*
7. *The derivative $\partial_\theta r_T(x,\theta)$ exists and is finite for any $x \in E$.*
8. *For any $x \in E$, $\mathrm{E}_\theta(\widehat{r}_T(x))^2 < \infty$ and $\partial_\theta \mathrm{E}_\theta[\widehat{r}_T(x)]$ exists and is differentiable under $\int$,* i.e.

$$\partial_\theta \mathrm{E}_\theta[\widehat{r}_T(x)] = \int_{\mathscr{X}_T} \partial_\theta \widehat{r}_T(x) f_T(\cdot,\theta)d\nu.$$

Remark 2.1. In condition 8 above, the integration is not performed with respect to x, but with respect to the first variable of f_T, this is a sample path of $X_{(T)}$, which the estimator $\widehat{r}_T$ depends on.

We denote by

$$b_T(x,\theta) = \mathrm{E}_\theta \widehat{r}_T(x) - r_T(x,\theta), \quad x \in E, \quad T \geqslant 0,$$

the bias of $\widehat{r}_T(x)$ for the estimation of $r_T(x,\theta)$.

Proposition 2.1. *Under Assumption 2.1,*

$$\rho_T(\theta) = \int_E \mathrm{E}_\theta\left(\widehat{r}_T(x) - r_T(x,\theta)\right)^2 d\mu_{\theta,T}(x) \geqslant \frac{\|\partial_\theta \mathrm{E}_\theta \widehat{r}_T\|^2_{\mu_{\theta,T}}}{I_T(\theta)} + \|b_T(\cdot,\theta)\|^2_{\mu_{\theta,T}}. \tag{1}$$

In particular if $\widehat{r}_T = r_T(\cdot,\widehat{\theta}_T)$ with $\widehat{\theta}_T$ an estimator of θ, then

$$\rho_T(\theta) = \int_E \mathrm{E}_\theta\left(r_T(x,\widehat{\theta}_T) - r_T(x,\theta)\right)^2 d\mu_{\theta,T}(x) \geqslant \frac{\|\partial_\theta \mathrm{E}_\theta r_T(\cdot,\widehat{\theta}_T)\|^2_{\mu_{\theta,T}}}{I_T(\theta)} + \|b_T(\cdot,\theta)\|^2_{\mu_{\theta,T}}. \tag{2}$$

Proof. Let $x \in E$. Applying the Cramér-Rao inequality (see [8] p. 120) we get

$$\mathrm{E}_\theta\left(\widehat{r}_T(x) - r_T(x,\theta)\right)^2 \geqslant \frac{\left(\partial_\theta \mathrm{E}_\theta \widehat{r}_T(x)\right)^2}{I_T(\theta)} + b_T(x,\theta)^2.$$

One integrates with respect to $\mu_{\theta,T}$ and obtains the desired result. □

Remark 2.2. The lower bound only depends on the estimator $\widehat{r}_T$ through its bias $b_T(\cdot,\theta)$ and the derivative of its bias with respect to θ, since $\mathrm{E}_\theta \widehat{r}_T = r_T(\cdot,\theta) + b_T(\cdot,\theta)$.

Remark 2.3. Obviously the same result holds when there is no dependence in T.

2.3. Asymptotic efficiency

In addition to Assumption 2.1 we do the following assumption on the bias of the estimator $\widehat{r}_T$.

Assumption 2.2.

$$\|\partial_\theta b_T(\cdot,\theta)\|_{\mu_{\theta,T}} = o(\|\partial_\theta r_T(\cdot,\theta)\|_{\mu_{\theta,T}}), \quad T\to\infty$$

We now state an asymptotic bound on the QER under Assumptions 2.1 and 2.2. We will use the following notation.

$$v_T(\theta) = \frac{\|\partial_\theta r_T(\cdot,\theta)\|^2_{\mu_{\theta,T}}}{I_T(\theta)}$$

Proposition 2.2. *Under Assumptions 2.1 and 2.2, for any $\theta\in\Theta$,*

$$\underline{\lim}_{T\to\infty} \frac{\rho_T(\theta)}{v_T(\theta)} \geqslant 1.$$

Proof. From the information inequality (1)

$$\rho_T(\theta) \geqslant \frac{\|\partial_\theta \mathrm{E}_\theta \widehat{r}_T\|^2_{\mu_{\theta,T}}}{I_T(\theta)} + \|b_T(\cdot,\theta)\|^2_{\mu_{\theta,T}} \geqslant \frac{\|\partial_\theta \mathrm{E}_\theta \widehat{r}_T\|^2_{\mu_{\theta,T}}}{I_T(\theta)}.$$

Hence

$$\frac{\rho_T(\theta)}{v_T(\theta)} \geqslant \frac{\|\partial_\theta \mathrm{E}_\theta \widehat{r}_T\|^2_{\mu_{\theta,T}}}{\|\partial_\theta r_T(\cdot,\theta)\|^2_{\mu_{\theta,T}}}. \tag{3}$$

Now

$$\partial_\theta \mathrm{E}_\theta \widehat{r}_T(x) = \partial_\theta r_T(x,\theta) + \partial_\theta b_T(x,\theta), \quad x\in E$$

hence

$$\frac{\|\partial_\theta \mathrm{E}_\theta \widehat{r}_T\|_{\mu_{\theta,T}}}{\|\partial_\theta r_T(\cdot,\theta)\|_{\mu_{\theta,T}}} \geqslant 1 - \frac{\|\partial_\theta b_T(\cdot,\theta)\|_{\mu_{\theta,T}}}{\|\partial_\theta r_T(\cdot,\theta)\|_{\mu_{\theta,T}}}.$$

Now using Assumption 2.2 we obtain

$$\underline{\lim}_{T\to\infty} \frac{\|\partial_\theta \mathrm{E}_\theta \widehat{r}_T\|_{\mu_{\theta,T}}}{\|\partial_\theta r_T(\cdot,\theta)\|_{\mu_{\theta,T}}} \geqslant 1. \tag{4}$$

Combining with (3) we obtain the result. □

Proposition 2.2 leads naturally to the following definition of asymptotic efficiency for estimators of the regression function.

Definition 2.1. Suppose the model satisfies Assumption 2.1 and let $\widehat{r}_T$ be an estimator of $r_T(\cdot,\theta)$ that satisfies Assumptions 2.1 and 2.2. Let $\mu_{\theta,T}$ be a measure on $(E,\mathscr{B})$ to define the QER. We will say that $\widehat{r}_T$ is an *asymptotically efficient* estimator of $r_T(\cdot,\theta)$ if

$$\rho_T(\theta) \sim v_T(\theta) = \frac{\|\partial_\theta r_T(\cdot,\theta)\|^2_{\mu_{\theta,T}}}{I_T(\theta)}, \quad T\to\infty, \quad \forall\theta\in\Theta.$$

Remark 2.4. While developping our results, we always have in mind the regression function for $r_T(\cdot,\theta)$, but all the results remain true for any other function satisfying the assumptions. In particular, taking $r_T(\cdot,\theta) \equiv \theta$, the constant function uniformly equals to θ, in the definition above and in the corresponding assumptions, yields a definition of asymptotic efficiency for estimators of the parameter θ.

Remark 2.5. Denote by

$$m_T(\theta) = \frac{\|\partial_\theta \mathrm{E}_\theta \widehat{r}_T\|^2_{\mu_{\theta,T}}}{I_T(\theta)} + \|b_T(\cdot,\theta)\|^2_{\mu_{\theta,T}},$$

the lower bound in (1). It might seem more natural to define asymptotic efficiency as $\rho_T(\theta) \sim m_T(\theta)$, $T \to \infty$, $\forall \theta \in \Theta$. However $m_T(\theta)$ depends on the estimator $\widehat{r}_T$ (through its bias and the derivative of its bias), while $v_T(\theta)$ does not depend at all on the estimator, which makes it more suitable for the definition of an optimality criteria. It is an advantage of asymptotic efficiency over the Cramér-Rao lower bound. Moreover we see in the following proposition that in case of asymptotic efficiency, $m_T(\theta) \sim v_T(\theta)$ holds.

Proposition 2.3. *Let $\widehat{r}_T$ be an asymptotically efficient estimator of $r_T(\cdot,\theta)$, and let $\theta \in \Theta$. Then, under Assumptions 2.1 and 2.2, as $T \to \infty$, the following conditions hold*

$$\|\partial_\theta \mathrm{E}_\theta \widehat{r}_T\|_{\mu_{\theta,T}} \sim \|\partial_\theta r_T(\cdot,\theta)\|_{\mu_{\theta,T}}, \tag{5}$$

$$\frac{\|b_T(\cdot,\theta)\|^2_{\mu_{\theta,T}}}{\|\partial_\theta r_T(\cdot,\theta)\|^2_{\mu_{\theta,T}}} = o(I_T^{-1}(\theta)), \tag{6}$$

$$v_T(\theta) \sim m_T(\theta) \sim \rho_T(\theta). \tag{7}$$

Proof. From

$$\partial_\theta \mathrm{E}_\theta \widehat{r}_T(x) = \partial_\theta r_T(x,\theta) + \partial_\theta b_T(x,\theta), \quad x \in E$$

we get

$$\frac{\|\partial_\theta \mathrm{E}_\theta \widehat{r}_T\|_{\mu_{\theta,T}}}{\|\partial_\theta r_T(\cdot,\theta)\|_{\mu_{\theta,T}}} \leqslant 1 + \frac{\|\partial_\theta b_T(\cdot,\theta)\|_{\mu_{\theta,T}}}{\|\partial_\theta r_T(\cdot,\theta)\|_{\mu_{\theta,T}}}.$$

thus Assumption 2.2 implies

$$\overline{\lim_{T\to\infty}} \frac{\|\partial_\theta \mathrm{E}_\theta \widehat{r}_T\|_{\mu_{\theta,T}}}{\|\partial_\theta r_T(\cdot,\theta)\|_{\mu_{\theta,T}}} \leqslant 1.$$

Combining with (4) we obtain (5).

On the other hand, from the information inequality (1)

$$\frac{\rho_T(\theta)}{v_T(\theta)} \geqslant \frac{\|\partial_\theta \mathrm{E}_\theta \widehat{r}_T\|^2_{\mu_{\theta,T}}}{\|\partial_\theta r_T(\cdot,\theta)\|^2_{\mu_{\theta,T}}} + \varepsilon_T$$

where

$$\varepsilon_T = \frac{\|b_T(\cdot,\theta)\|^2_{\mu_{\theta,T}}}{I_T^{-1}(\theta)\|\partial_\theta r_T(\cdot,\theta)\|^2_{\mu_{\theta,T}}}$$

therefore the condition $\frac{\rho_T(\theta)}{v_T(\theta)} \to 1$ and (5) imply $\overline{\lim}\, \varepsilon_T \leqslant 0$, hence (6).

Finally with (5) and (6)

$$\frac{m_T(\theta)}{v_T(\theta)} = \frac{\|\partial_\theta \mathrm{E}_\theta \widehat{r}_T\|^2_{\mu_{\theta,T}}}{\|\partial_\theta r_T(\cdot,\theta)\|^2_{\mu_{\theta,T}}} + \varepsilon_T \xrightarrow[T\to\infty]{} 1.$$

□

Example 2.1. *Ornstein-Uhlenbeck process*

We resume with the problem of prediction of a stationary Ornstein-Uhlenbeck process as stated in the introduction. In this problem the conditional expectation of interest is $\mathrm{E}_\theta[X_{T+h}|X_T] = X_T e^{-\theta h}$. We consider the associated problem of estimation of the regression function $x \mapsto r_T(x,\theta) = xe^{-\theta h}$ and the plug-in estimator $r_T(x,\widehat{\theta}_T) = xe^{-\widehat{\theta}_T h}$, where $\widehat{\theta}_T$ is the maximum likelihood estimator (MLE) of θ based on $X_{(T)}$. We choose $\mu_{\theta,T}$ as the distribution of X_T, *i.e.* $\mu_{\theta,T} = \mathscr{N}(0,\frac{1}{2\theta})$.

The Fisher information associated with $X_{(T)}$ is (see [6] p. 55)

$$I_T(\theta) = I_0(\theta) + TI(\theta) \quad \text{with} \quad I(\theta) = \frac{1}{2\theta}.$$

Let us see that Assumption 2.1 is satisfied. All the distributions of the model are mutally absolutely continuous and the same is true for the corresponding distributions of $X_{(T)}$. Hence we can choose $\nu = \mathrm{P}_{\theta_0}$ for some fixed $\theta_0 \in \mathbb{R}_+^*$. Then the density of $X_{(T)}$ with respect to P_{θ_0} is the function $f_T(\cdot,\theta)$ such that (see [6] p. 37–38)

$$f_T(X_{(T)},\theta) = \sqrt{\frac{\theta}{\theta_0}} \exp\left\{\frac{\theta-\theta_0}{2}(T - X_0^2 - X_T^2) - \frac{\theta^2-\theta_0^2}{2}\int_0^T X_t^2 dt\right\}.$$

The function $b_T(\cdot,\theta)$ is

$$b_T(x,\theta) = x\left(\mathrm{E}_\theta\left(e^{-\widehat{\theta}_T h}\right) - e^{-\theta h}\right).$$

It is straightforward to see that Assumption 2.1 is fulfilled.

[2] proved that $\mathrm{E}_\theta\left(e^{-\widehat{\theta}_T h}\right) - e^{-\theta h} \xrightarrow[T\to\infty]{} 0$. Hence $\|\partial_\theta b_T(\cdot,\theta)\|^2_{\mu_{\theta,T}} \longrightarrow 0$. Now

$$\|\partial_\theta r_T(\cdot,\theta)\|_{\mu_{\theta,T}} = \int_E (-he^{-\theta h})^2 x^2 d\mu_{\theta,T}(x) = \frac{h^2 e^{-2\theta h}}{2\theta},$$

which does not depend on T. We deduce that Assumption 2.2 is fulfilled.

Now we are going to see that the plug-in estimator $r_T(x,\widehat{\theta}_T) = xe^{-\widehat{\theta}_T h}$, where $\widehat{\theta}_T$ is the MLE of θ, is asymptotically efficient for estimating the regression function $x \mapsto xe^{-\theta h}$. First

$$v_T(\theta) = \frac{\|\partial_\theta r_T(\cdot,\theta)\|^2_{\mu_{\theta,T}}}{I_T(\theta)} = \frac{h^2 e^{-2\theta h}}{T + 2\theta I_0(\theta)}$$

thus the condition of asymptotic efficiency is

$$\rho_T(\theta) = \mathrm{E}_\theta(e^{-\widehat{\theta}_T h} - e^{-\theta h})^2 \int_E x^2 d\mu_{\theta,T} \sim \frac{h^2 e^{-2\theta h}}{T}$$

i.e.

$$\mathrm{E}_\theta(e^{-\widehat{\theta}_T h} - e^{-\theta h})^2 \sim \frac{2\theta h^2 e^{-2\theta h}}{T}.$$

Now if $\eta = e^{-\theta h}$, then the MLE of η is $\widehat{\eta}_T = e^{-\widehat{\theta}_T h}$ and the following holds (see [6] p. 121)

$$\lim_{T\to\infty} T\mathrm{E}_\eta(\widehat{\eta}_T - \eta)^2 = \lim_{T\to\infty} \frac{T}{I_T(\eta)} = I^{-1}(\eta)$$

with an abuse of notation, $I_T(\eta)$ and $I(\eta)$ stand for the Fisher information and the asymptotic Fisher information when η is the parameter of the model.

The formula of change of parameterization for the Fisher information ([8] p. 125) gives

$$I(\theta) = I(\eta)h^2 e^{-2\theta h}.$$

Hence

$$T\mathrm{E}_\eta(\widehat{\eta}_T - \eta)^2 \longrightarrow 2\theta h^2 e^{-2\theta h},$$

and therefore

$$\mathrm{E}_\theta(e^{-\widehat{\theta}_T h} - e^{-\theta h})^2 \sim \frac{2\theta h^2 e^{-2\theta h}}{T},$$

which shows the asymptotic efficiency of the estimator $\widehat{r}_T : x \mapsto x e^{-\widehat{\theta}_T h}$ for estimating the regression function $x \mapsto x e^{-\theta h}$. In particular Proposition 2.3 applies.

2.4. *Conditions for asymptotic efficiency*

From now on we are interested in plug-in estimators of the form $\widehat{r}_T = r_T(\cdot, \widehat{\theta}_T)$ where $\widehat{\theta}_T$ is an estimator of θ based on $X_{(T)}$.

We first remark that under a simple (but restrictive) condition the estimator of the regression function $r_T(\cdot, \widehat{\theta}_T)$ and the predictor $r_T(X_T, \widehat{\theta}_T)$ converge at least at the same rate than the associated estimator $\widehat{\theta}_T$ of the parameter θ.

Proposition 2.4. *If $r_T(x,\theta)$ is Lipschitz with respect to θ, uniformly in x,* i.e. *there exists $c > 0$ such that $\forall \theta, \theta' \in \Theta$*

$$\sup_{x\in E} \left| r_T(x,\theta') - r_T(x,\theta) \right| \leqslant c|\theta' - \theta|,$$

then for any estimator $\widehat{\theta}_T$ of θ, the QER of $r_T(\cdot, \widehat{\theta}_T)$ and the QEP of $r_T(X_T, \widehat{\theta}_T)$ are $\mathscr{O}\big(\mathrm{E}_\theta(\widehat{\theta}_T - \theta)^2\big)$.

Proof. The Lipschitz condition gives

$$\int_E \mathrm{E}_\theta \left(r_T(x,\widehat{\theta}_T) - r_T(x,\theta) \right)^2 d\mu_{\theta,T}(x) \leqslant \mathrm{E}_\theta \left(\sup_{x\in E} \left| r_T(x,\widehat{\theta}_T) - r_T(x,\theta) \right| \right)^2 \leqslant c^2 \mathrm{E}_\theta(\widehat{\theta}_T - \theta)^2,$$

and

$$\mathrm{E}_\theta \left(r_T(X,\widehat{\theta}_T) - r_T(X,\theta) \right)^2 \leqslant \mathrm{E}_\theta \left(\sup_{x\in E} |r_T(x,\widehat{\theta}_T) - r_T(x,\theta)| \right)^2 \leqslant c^2 \mathrm{E}_\theta(\widehat{\theta}_T - \theta)^2.$$

The result follows. □

For a sharper result about the QER we use the following assumptions.

Assumption 2.3.

1. *$\partial_\theta r_T(x,\theta)$ exists $\forall x \in E$, $\forall \theta \in \Theta$, $\forall T > 0$.*
2. *$r_T(\cdot,\theta) \in \mathrm{L}^2(\mu_{\theta,T})$ and $\|\partial_\theta r_T(\cdot,\theta)\|^2_{\mu_{\theta,T}} \xrightarrow[T\to\infty]{} U(\theta)$, for all $\theta \in \Theta$,*
3. *$\exists \alpha \in (0,1]$, $\exists c_T(x)$ such that $\sup_T \|c_T\|_{\mu_{\theta,T}} < \infty$ and*

$$|\partial_\theta r_T(x,\theta') - \partial_\theta r_T(x,\theta)| \leqslant c_T(x)|\theta' - \theta|^\alpha, \quad \forall x \in E, \quad \forall \theta, \theta' \in \Theta, \quad \forall T > 0.$$

In addition, $\widehat{\theta}_T$ is an estimator of θ with values in Θ such that

4. *$\lim_{T\to\infty} T\mathrm{E}_\theta(\widehat{\theta}_T - \theta)^2 = V(\theta) \in (0,\infty)$,*
5. *$\mathrm{E}_\theta|\widehat{\theta}_T - \theta|^{2+2\alpha} = o\left(\frac{1}{T}\right)$.*

Proposition 2.5. *Under Assumption 2.3,*

$$T\rho_T(\theta) \xrightarrow[T\to\infty]{} U(\theta)V(\theta).$$

Proof. We have

$$r_T(x,\widehat{\theta}_T) - r_T(x,\theta) = (\widehat{\theta}_T - \theta)\partial_\theta r_T(x,\widetilde{\theta}_T)$$

where $\widetilde{\theta}_T \in [\min(\theta,\widehat{\theta}_T), \max(\theta,\widehat{\theta}_T)]$. Let

$$\delta_T(x) = (\widehat{\theta}_T - \theta)(\partial_\theta r_T(x,\widetilde{\theta}_T) - \partial_\theta r_T(x,\theta)).$$

Hence

$$r_T(x,\widehat{\theta}_T) - r_T(x,\theta) = (\widehat{\theta}_T - \theta)\partial_\theta r_T(x,\theta) + \delta_T(x),$$

and using condition 3 of Assumption 2.3

$$|\delta_T(x)| \leqslant c_T(x)|\widehat{\theta}_T - \theta|^{1+\alpha}. \tag{8}$$

On the other hand,

$$\begin{aligned}
\rho_T(\theta) &= \mathrm{E}_\theta(\widehat{\theta}_T - \theta)^2 \int_E (\partial_\theta r_T(x,\theta))^2 d\mu_{\theta,T}(x) + \int_E \mathrm{E}_\theta(\delta_T^2(x))d\mu_{\theta,T}(x) \\
&\quad + 2\int_E \mathrm{E}_\theta[(\widehat{\theta}_T - \theta)\delta_T(x)]\partial_\theta r_T(x,\theta)d\mu_{\theta,T}(x) \\
&= J_1 + J_2 + J_3.
\end{aligned}$$

From (8) it follows that

$$|J_2| \leqslant \mathrm{E}_\theta(|\widehat{\theta}_T - \theta|^{2+2\alpha}) \int_E c_T^2 d\mu_{\theta,T}.$$

Since $\|c_T\|_{\mu_{\theta,T}}$ is bounded, condition 5 of Assumption 2.3 implies $|J_2| = o(\frac{1}{T})$. Now (8) gives

$$\begin{aligned}
|J_3| &\leqslant 2\mathrm{E}_\theta(|\widehat{\theta}_T - \theta|^{2+\alpha}) \int_E c_T |\partial_\theta r_T(\cdot,\theta)| d\mu_{\theta,T} \\
&\leqslant 2\mathrm{E}_\theta|\widehat{\theta}_T - \theta|^{2+\alpha} \|c_T\|_{\mu_{\theta,T}} \|\partial_\theta r_T(\cdot,\theta)\|_{\mu_{\theta,T}} \\
&\leqslant 2\mathrm{E}_\theta \left[|\widehat{\theta}_T - \theta||\widehat{\theta}_T - \theta|^{1+\alpha} \right] \|c_T\|_{\mu_{\theta,T}} \|\partial_\theta r_T(\cdot,\theta)\|_{\mu_{\theta,T}} \\
&\leqslant 2 \left(\mathrm{E}_\theta(\widehat{\theta}_T - \theta)^2 \right)^{\frac{1}{2}} \left(\mathrm{E}_\theta|\widehat{\theta}_T - \theta|^{2+2\alpha} \right)^{\frac{1}{2}} \|c_T\|_{\mu_{\theta,T}} \|\partial_\theta r_T(\cdot,\theta)\|_{\mu_{\theta,T}} \\
&\leqslant \mathcal{O}\left(\frac{1}{\sqrt{T}} \right) o\left(\frac{1}{\sqrt{T}} \right) = o\left(\frac{1}{T} \right).
\end{aligned}$$

Finally $TJ_1 \longrightarrow U(\theta)V(\theta)$. □

Remark 2.6. Under Assumptions 2.1, 2.2 and 2.3, the estimator $r_T(\cdot,\widehat{\theta}_T)$ is asymptotically efficient for the QER as soon as $V(\theta) = \lim_{T\to\infty} \frac{T}{I_T(\theta)}$.

Example 2.2. *Ornstein-Uhlenbeck process*

Here we choose $\mu_{\theta,T} = \mathrm{P}_{\theta,X_T} = \mathrm{P}_{\theta,X_0}$, the distribution of X_T under P_θ. The maximum likelihood estimator (MLE) $\widehat{\theta}_T$ satisfies conditions 4 and 5 of Assumption 2.3 with $V(\theta) = I(\theta)^{-1} = 2\theta$ (see [6] p. 121).

On the other hand $\partial_\theta r_T(x,\theta) = -hxe^{-\theta h}$ hence

$$\mathrm{E}_\theta(\partial r_T(X_T,\theta))^2 = h^2 e^{-2\theta h}\mathrm{E}_\theta(X_T^2) = \frac{h^2 e^{-2\theta h}}{2\theta} \in (0,\infty).$$

and

$$|\partial_\theta r_T(x,\theta') - \partial_\theta r_T(x,\theta)| = (\partial_\theta)^2 r_T(x,\theta'')|\theta' - \theta|, \quad \theta'' \in [\min(\theta,\theta'),\max(\theta,\theta')].$$

with $|(\partial_\theta)^2 r_T(x,\theta'')| = |h^2 x e^{-\theta'' h}| \leqslant h^2|x|$ thus

$$\sup_T \int_E ((\partial_\theta)^2 r_T(\cdot,\theta))^2 dP_{\theta,X_0} \leqslant h^4 \int_E x^2 dP_{\theta,X_0}(x) = \frac{h^4}{2\theta} < \infty,$$

Therefore we can apply Proposition 2.5 with $U(\theta) = \frac{h^2 e^{-2\theta h}}{2\theta}$ and $V(\theta) = 2\theta$, and we find $T\rho_T(\theta) \longrightarrow h^2 e^{-2\theta h}$.

Remark that instead of choosing $\mu_{\theta,T} = \mathrm{P}_{\theta,X_T}$, we might alternatively choose $\mu_{\theta,T} = \delta_{(x_0)}$ with $x_0 \in \mathbb{R}$. This would lead to results about estimation of the regression function evaluated at one particular point x_0.

3. Risk comparison

In this section we give conditions under which the QEP and the QER are asymptotically equivalent. This will allow to define asymptotic efficiency for predictors and make possible to use the results of section 2 for prediction problems. As in section 2, the process $(X_t, t \geqslant 0)$ is assumed to be Markov. The QER will always be taken with respect to $\mu_{\theta,T} = P_{\theta,X_T}$ the distribution of X_T.

3.1. Assumptions and lemmas

We are now interested in the QEP defined by

$$R_T(\theta) = \mathrm{E}_\theta\left(r_T(X_T,\widehat{\theta}_T) - r_T(X_T,\theta)\right)^2, \quad \theta \in \Theta.$$

In order to compare it with $\rho_T(\theta)$ we consider an auxiliary predictor $r_T(X_T,\widehat{\theta}_{S(T)})$ and the corresponding estimator of the regression function $r_T(\cdot,\widehat{\theta}_{S(T)})$ where $\widehat{\theta}_{S(T)}$ is based on $X_{(S(T))}$ with $S:\mathbb{R}_+ \to \mathbb{R}_+$ an increasing function such that $S(T) < T$ for all $T \geqslant 0$ and $S(T) \sim T$, as $T \to \infty$. In what follows, S will always stand for $S(T)$, omitting the argument T to make notations lighter.

For all $T \geqslant 0$, we define the function $\bar{\theta}_T : \mathscr{X}_T \to \Theta$, such that $\bar{\theta}_T(X_{(T)}) = \widehat{\theta}_T$.

Let

$$\Delta r(x,\xi) = r_T(x,\bar{\theta}_S(\xi)) - r_T(x,\theta), \quad \forall x \in E, \xi \in \mathscr{X}_S,$$

where the dependence of Δr in T and θ is kept implicit, to make notations lighter.

The distribution $\mathrm{P}_{\theta,(X_T,X_{(S)})}$ of $(X_T,X_{(S)})$, is assumed to be dominated by a σ-finite measure λ and f_{θ,X_T}, $f_{\theta,X_{(S)}}$ and $f_{\theta,(X_T,X_{(S)})}$ stand for the densities of X_T, $X_{(S)}$ and $(X_T,X_{(S)})$ respectively. Now let

$$R_T^S(\theta) = \int_{E\times\mathscr{X}_S} (\Delta r(x,\xi))^2 dP_{\theta,(X_T,X_{(S)})}(x,\xi) = \int_{E\times\mathscr{X}_S} (\Delta r(x,\xi))^2 f_{\theta,(X_T,X_{(S)})} d\lambda(x,\xi)$$

and

$$\rho_T^S(\theta) = \int_{E\times\mathscr{X}_S} (\Delta r(x,\xi))^2 dP_{\theta,X_T}(x) dP_{\theta,X_{(S)}}(\xi) = \int_{E\times\mathscr{X}_S} (\Delta r(x,\xi))^2 f_{\theta,X_T}(x) f_{\theta,X_{(S)}}(\xi) d\lambda(x,\xi)$$

Also let

$$\Delta f(x,\xi) = |f_{\theta,(X_T,X_{(S)})}(x,\xi) - f_{\theta,X_T}(x) f_{\theta,X_{(S)}}(\xi)|,$$

keeping the dependence of Δf in T and θ implicit.

We measure the dependence of X_T and $X_{(S)}$ by the coefficient

$$\widetilde{\beta}(S,T) = \int_{E\times\mathscr{X}_S} (\Delta f) d\lambda.$$

Remark 3.1. It is bounded from above by the usual β-mixing coefficient. It holds $\widetilde{\beta}(S,T) \leqslant 2\beta(T-S)$ with

$$\beta(t) = \sup_{s>0} \left\| \mathrm{P}_{0,\theta}^s \otimes \mathrm{P}_{s+t,\theta}^\infty - \mathrm{P}_{s,t,\theta} \right\|_{\mathrm{TV}}$$

where $\mathrm{P}_{0,\theta}^s$ is the distribution of $(X_u)_{0\leqslant u\leqslant s}$, $\mathrm{P}_{s+t,\theta}^\infty$ the distribution of $(X_u)_{u\geqslant s+t}$, $\mathrm{P}_{s,t,\theta}$ the joint distribution of $((X_u)_{0\leqslant u\leqslant s},(X_u)_{u\geqslant s+t})$, and $\|\cdot\|_{\mathrm{TV}}$ the total variation norm for signed measures, *i.e.* if μ is a signed measure on a measurable space $\mathscr{A}$, then $\|\mu\|_{\mathrm{TV}} = \sup_{A\in\mathscr{A}} |\mu(A)|$. For a reference about mixing coefficients see [4].

We now make the following assumption.

Assumption 3.1.

1. $\exists m \in (2,\infty] : \delta_{\theta,m} = \sup_{S,T} \|\Delta r\|_{\mathrm{L}^m(\mu)} < \infty$, *with* $\mu = (\Delta f)\lambda$,
2. $\widetilde{\beta}(S,T) = o\left((\rho_T^S)^{\frac{m}{m-2}}\right)$, *as* $T \to \infty$.

Lemma 3.1. *Under Assumption 3.1,* $R_T^S \sim \rho_T^S$ *as* $T \to \infty$.

Proof. Suppose $m < \infty$, we have

$$|R_T^S - \rho_T^S| \leqslant \int (\Delta r)^2 \Delta f d\lambda.$$

Applying Hölder inequality to $(\Delta r)^2 (\Delta f)^{\frac{1}{p}}$ and $(\Delta f)^{\frac{1}{q}}$ with $p = \frac{m}{2}$ and $q = \frac{m}{m-2}$ we get

$$|R_T^S - \rho_T^S| \leqslant \left(\int (\Delta r)^m \Delta f d\lambda\right)^{\frac{2}{m}} \left(\int \Delta f d\lambda\right)^{\frac{m-2}{m}} \leqslant \delta_{\theta,m}^2 \widetilde{\beta}(S,T)^{1-\frac{2}{m}}.$$

It is straightforward to verify this inequality for the case $m = \infty$. We deduce

$$\left|\frac{R_T^S}{\rho_T^S} - 1\right| \leqslant \delta_{\theta,m}^2 \frac{\widetilde{\beta}(S,T)^{1-\frac{2}{m}}}{\rho_T^S} \longrightarrow 0.$$

□

Example 3.1. *Ornstein-Uhlenbeck process*
We take $S = T - \sqrt{T}$. Here $\lambda = \ell \otimes \mathrm{P}_{\theta_0, X_{(S)}}$ with ℓ the Lebesgue measure on $\mathbb{R}$ and $\mathrm{P}_{\theta_0, X_{(S)}}$ the distribution of $X_{(S)}$ under some fixed parameter $\theta_0 \in \Theta$. It holds $|\Delta r(x,\xi)| \leqslant |x|$, for all $x \in \mathbb{R}$. We deduce the first condition of Assumption 3.1 is fulfilled,

$$\|\Delta r\|_{\mathrm{L}^m(\mu)}^m \leqslant \int |\Delta r|^m \Delta f d\lambda \leqslant 2\mathrm{E}_\theta |X_0|^m < \infty.$$

In particular for $m = 4$, this upper bound is $2\mathrm{E}_\theta |X_0|^4 = 6(\mathrm{E}_\theta X_0^2)^2 = \frac{3}{\theta}$. For the second condition, we have

$$\rho_T^S = (\mathrm{E}_\theta X_T^2)\mathrm{E}_\theta \left(e^{-\widehat{\theta}_S h} - e^{-\theta h}\right)^2 \sim \frac{C}{T}, \text{ as } T \to \infty,$$

for some constant C (see [6]). The Ornstein-Uhlenbeck process is geometrically β-mixing (see [13]), *i.e.* there exists some $\gamma > 0$ such that $\beta(t) \leqslant e^{-\gamma t}$. Moreover

$$\widetilde{\beta}(S,T) \leqslant 2\beta(T-S) = 2\beta(\sqrt{T}).$$

Hence the second condition of Assumption 3.1 is fulfilled too for any $m > 2$.

We are now looking for conditions under which $R_T \sim \rho_T$. We will use the following lemma.

Lemma 3.2. *Let* (F,d) *be a metric space,* (u_n) *a sequence of real numbers, then the conditions* $u_n d(a_n,b_n) \to l$ *and* $u_n d(b_n,c_n) \to 0$ *imply* $u_n d(a_n,c_n) \to l$.

We denote by

$$b_{\widehat{\theta}_T}(\theta) = \mathrm{E}_\theta \widehat{\theta}_T - \theta$$

the bias of the estimator $\widehat{\theta}_T$. And we make the following assumption.

Assumption 3.2. *Condition 2 of Assumption 2.1 holds and,*

1. *for any $x \in E$,* $\mathrm{E}_\theta(\widehat{\theta}_T)^2 < \infty$ *and* $\partial_\theta \mathrm{E}_\theta \widehat{\theta}_T$ *exists and is differentiable under* $\int$, i.e.

$$\partial_\theta \mathrm{E}_\theta \widehat{\theta}_T = \int_{\mathscr{X}_T} \partial_\theta \left(\bar{\theta}_T(\xi) f_T(\xi,\theta)\right) d\nu(\xi).$$

In addition, as $T \to \infty$, the following conditions hold.

2. *There exists $I(\theta) > 0$ such that $I_T(\theta) \sim TI(\theta)$.*
3. $T\mathrm{E}_\theta(\widehat{\theta}_T - \theta)^2 \to I(\theta)^{-1}$.
4. $\partial_\theta b_{\widehat{\theta}_T}(\theta) \to 0$.

Remark 3.2. Under Assumptions 2.1 and 3.2, the estimator $\widehat{\theta}_T$ fulfills the definition of asymptotic efficiency of Remark 2.4. Conversely, if $I_T(\theta) \sim TI(\theta)$ and $\widehat{\theta}_T$ is an estimator of θ that fulfills the definition of asymptotic efficiency of Remark 2.4, then Assumption 3.2 is satisfied.

Lemma 3.3. *Under Assumption 3.2,*

$$T\mathrm{E}_\theta(\widehat{\theta}_T - \widehat{\theta}_S)^2 \xrightarrow[T\to\infty]{} 0.$$

Proof. Let $\widetilde{\theta}_T = \frac{1}{2}(\widehat{\theta}_T + \widehat{\theta}_S)$. For any vectors x and y, the parallelogram identity is

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2.$$

Taking $x = \widehat{\theta}_T - \theta$ and $y = \widehat{\theta}_S - \theta$ in the normed space $\mathrm{L}^2(\mathrm{P}_\theta)$ one gets

$$\mathrm{E}_\theta(\widehat{\theta}_T + \widehat{\theta}_S - 2\theta)^2 + \mathrm{E}_\theta(\widehat{\theta}_T - \widehat{\theta}_S)^2 = 2E_\theta(\widehat{\theta}_T - \theta)^2 + 2\mathrm{E}_\theta(\widehat{\theta}_S - \theta)^2.$$

Hence

$$4T\mathrm{E}_\theta(\widetilde{\theta}_T - \theta)^2 + T\mathrm{E}_\theta(\widehat{\theta}_T - \widehat{\theta}_S)^2 = 2T\mathrm{E}_\theta(\widehat{\theta}_T - \theta)^2 + 2T\mathrm{E}_\theta(\widehat{\theta}_S - \theta)^2.$$

Now to complete the proof it remains to prove that

$$T\mathrm{E}_\theta(\widetilde{\theta}_T - \theta)^2 \xrightarrow[T\to\infty]{} I(\theta)^{-1}. \tag{9}$$

From

$$|a+b|^2 \leqslant 2|a|^2 + 2|b|^2,$$

with $a = \frac{\widehat{\theta}_T - \theta}{2}$ and $a = \frac{\widehat{\theta}_S - \theta}{2}$ one obtains

$$T\mathrm{E}_\theta(\widetilde{\theta}_T - \theta)^2 \leqslant 2\left[\frac{1}{4}T\mathrm{E}_\theta(\widehat{\theta}_T - \theta)^2 + \frac{1}{4}T\mathrm{E}_\theta(\widehat{\theta}_S - \theta)^2\right],$$

and since $S \sim T$,

$$\overline{\lim_{T\to\infty}}\, T\mathrm{E}_\theta(\widetilde{\theta}_T - \theta)^2 \leqslant 2\left[\frac{1}{4}I(\theta)^{-1} + \frac{1}{4}I(\theta)^{-1}\right] = I(\theta)^{-1}. \tag{10}$$

Now the information inequality gives

$$T\mathrm{E}_\theta(\widetilde{\theta}_T-\theta)^2 \geqslant T(b_{\widetilde{\theta}_T}(\theta))^2 + T\frac{\left(\partial_\theta b_{\widetilde{\theta}_T}(\theta)+1\right)^2}{I_T(\theta)},$$

and

$$\partial_\theta b_{\widetilde{\theta}_T}(\theta) = \frac{1}{2}\left[\partial_\theta b_{\widehat{\theta}_T}(\theta)+\partial_\theta b_{\widehat{\theta}_S}(\theta)\right] \xrightarrow[T\to\infty]{} 0,$$

therefore

$$\underline{\lim}_{T\to\infty} T\mathrm{E}_\theta(\widetilde{\theta}_T-\theta)^2 \geqslant \underline{\lim}_{T\to\infty} T\partial_\theta b_{\widetilde{\theta}_T}(\theta)+I(\theta)^{-1}. \quad (11)$$

From inequalities (10) and (11) we get $\underline{\lim}_{T\to\infty} T\partial_\theta b_{\widetilde{\theta}_T}(\theta)=0$. We deduce that (9) is true. This completes the proof. □

3.2. Asymptotic equivalence of the QER and the QEP

Assumption 3.3. *As* $T\to\infty$, $T\mathrm{E}_\theta(\widehat{\theta}_T-\widehat{\theta}_S)^2\to 0$.

Remark 3.3. Assumption 3.2 implies Assumption 3.3. Hence, if $I_T(\theta)\sim TI(\theta)$, and $\widehat{\theta}_T$ is an estimator of θ that fulfills the definition of asymptotic efficiency of Remark 2.4, then, by Lemma 3.3, Assumption 3.3 is satisfied.

Proposition 3.1. *Under Assumptions 2.3, 3.1 and 3.3, if there exists some deterministic constant* $C>0$ *such that*

$$|r_T(X_T,\theta^*)-r_T(X_T,\theta)| \leqslant C|\theta^*-\theta|, \quad \forall\theta,\theta^*\in\Theta, \quad (12)$$

then

$$R_T\sim\rho_T,\ \textit{as}\ T\to\infty.$$

Proof. With Assumption 3.1, Lemma 3.1 gives $R_T^S\sim\rho_T^S$. On the other hand, with Assumptions 2.3 and the condition $S\sim T$, Proposition 2.5 gives $\rho_T^S\sim\rho_T$. Now by condition (12)

$$T\mathrm{E}_\theta|r_T(X_T,\widehat{\theta}_T)-r_T(X_T,\widehat{\theta}_S)|^2 \leqslant C^2 T\mathrm{E}_\theta|\widehat{\theta}_T-\widehat{\theta}_S|^2.$$

With Assumption 3.3, the right-hand side vanishes as $T\to\infty$. Hence

$$T\mathrm{E}_\theta|r_T(X_T,\widehat{\theta}_T)-r_T(X_T,\widehat{\theta}_S)|^2 \xrightarrow[T\to\infty]{} 0.$$

Since

$$TR_T^S = T\mathrm{E}_\theta|r_T(X_T,\widehat{\theta}_S)-r_T(X_T,\theta)|^2 \longrightarrow U(\theta)V(\theta),$$

by Lemma 3.2 we deduce $R_T\sim R_T^S$, and finally,

$$R_T\sim R_T^S\sim\rho_T^S\sim\rho_T.$$

□

Proposition 3.2. *Suppose the following conditions hold.*

1. $T^2\mathrm{E}_\theta|\widehat{\theta}_T-\widehat{\theta}_S|^4=\mathscr{O}(1)$, *as* $T\to\infty$.

2. *There exists a measurable function $\ell : E \to \mathbb{R}_+$ such that*

$$|r_T(X_T,\theta^*) - r_T(X_T,\theta)| \leqslant \ell(X_T)|\theta^* - \theta|.$$

3. $C_\theta = \sup_T \mathrm{E}_\theta\left(\ell^{4+\nu}(X_T)\right) < \infty$ *for some $\nu > 0$.*

Then again, under Assumptions 2.3, 3.1 and 3.3,

$$R_T \sim \rho_T, \text{ as } T \to \infty.$$

Proof. Let $c > 0$ and

$$\Delta_T = T\mathrm{E}_\theta\big(r_T(X_T,\widehat{\theta}_T) - r_T(X_T,\widehat{\theta}_S)\big)^2 \leqslant T\mathrm{E}_\theta\left[\ell^2(X_T)(\widehat{\theta}_T - \widehat{\theta}_S)^2(\mathbb{1}_{\ell(X_T)\leqslant c} + \mathbb{1}_{\ell(X_T)>c})\right].$$

First, by Assumption 3.3

$$T\mathrm{E}_\theta\left(\ell^2(X_T)(\widehat{\theta}_T - \widehat{\theta}_S)^2\mathbb{1}_{\ell(X_T)\leqslant c}\right) \leqslant c^2 T\mathrm{E}_\theta(\widehat{\theta}_T - \widehat{\theta}_S)^2 \longrightarrow 0.$$

On the other hand, applying Hölder inequality with $p_1 = 1 + \frac{\nu}{8+\nu}$ and $q_1 = 1 + \frac{8+\nu}{\nu}$,

$$A_T = \mathrm{E}_\theta\left(\ell^2(X_T)(\widehat{\theta}_T - \widehat{\theta}_S)^2\mathbb{1}_{\ell(X_T)>c}\right) \leqslant \left(\mathrm{E}_\theta\left(\ell^{2p_1}(X_T)(\widehat{\theta}_T - \widehat{\theta}_S)^{2p_1}\right)\right)^{\frac{1}{p_1}} \left(\mathrm{P}_\theta(\ell(X_T) > c)\right)^{\frac{1}{q_1}}.$$

Applying Hölder inequality again with $p_2 = 2 + \frac{\nu}{4}$ and $q_2 = 1 + \frac{4}{4+\nu}$,

$$\begin{aligned}\left(\mathrm{E}_\theta\left(\ell^{2p_1}(X_T)(\widehat{\theta}_T - \widehat{\theta}_S)^{2p_1}\right)\right)^{\frac{1}{p_1}} &\leqslant \left(\mathrm{E}_\theta \ell^{2p_1p_2}(X_T)\right)^{\frac{1}{p_1p_2}} \left(\mathrm{E}_\theta(\widehat{\theta}_T - \widehat{\theta}_S)^{2p_1q_2}\right)^{\frac{1}{p_1q_2}} \\ &\leqslant C_\theta^{\frac{2}{4+\nu}} \left(\mathrm{E}_\theta(\widehat{\theta}_T - \widehat{\theta}_S)^4\right)^{\frac{1}{2}}.\end{aligned}$$

Since $2p_1p_2 = 4+\nu$ and $2p_1q_2 = 4$. Now, using the condition 1, there are some constants $M > 0$ and $T_0 > 0$, that do not depend on c, such that,

$$\sup_{T>T_0} TA_T \leqslant Mc^{-\frac{1}{q_1}}.$$

Now let $\varepsilon > 0$ and choose c such that $\sup_{T>T_0} TA_T < \varepsilon$. Hence $\overline{\lim}_{T\to\infty}\Delta_T < \varepsilon$. Therefore $\Delta_T \longrightarrow 0$. □

In a first version of this article, we used the condition $\mathrm{E}_\theta\ell^8(X_T) < \infty$ instead of the condition 3 of Proposition 3.2. We thank the anonymous referee for pointing us how the use of the Hölder inequality could weaken this condition.

It can be seen that the result applies to the Ornstein-Uhlenbeck process.

In light of Propositions 3.1 and 3.2 and section 2 we see that the following definition of asymptotic efficiency for plug-in predictors makes sense.

Definition 3.1. Let $r_T(X_T,\widehat{\theta}_T)$ be a plug-in predictor of $r_T(X_T,\theta)$ and suppose either the assumptions of Proposition 3.1 or those of Proposition 3.2 are fulfilled as well as Assumptions 2.1 and 2.2. We say that the predictor $r_T(X_T,\widehat{\theta}_T)$ is asymptotically efficient if

$$T\mathrm{E}_\theta(\widehat{\theta}_T - \theta)^2 \xrightarrow[T\to\infty]{} I(\theta)^{-1}.$$

4. Limit in distribution for plug-in predictors

In this section we weaken the assumption that the process $(X_t)_{t\geqslant 0}$ is Markov and assume instead

$$\mathrm{E}_\theta[X_{T+h}|X_{(T)}] = r_T(Z_T,\theta), \quad \forall T \geqslant 0,$$

where Z_T is $\sigma\big(X_t, \phi(T)\leqslant t\leqslant T\big)$-measurable, with $\frac{\phi(T)}{T}\to 1$, and $h>0$ is the horizon of prediction. We consider a plug-in predictor $r_T(Z_T,\widehat{\theta}_T)$ where $\widehat{\theta}_T$ is an estimator of θ.

4.1. Limit in distribution

We make the following assumptions.

Assumption 4.1.

1. *The process is α-mixing* i.e.

$$\alpha_\theta(u) = \sup_{t\geqslant 0}\ \sup_{\substack{A\in\mathscr{F}_0^t\\ B\in\mathscr{F}_{t+u}^\infty}} |\mathrm{P}_\theta(A\cap B)-\mathrm{P}_\theta(A)\mathrm{P}_\theta(B)| \xrightarrow[u\to\infty]{} 0$$

with $\mathscr{F}_0^t=\sigma(X_s, 0\leqslant s\leqslant t)$ and $\mathscr{F}_{t+u}^\infty=\sigma(X_s, t+u\leqslant s<\infty)$.

2. *There exist two independent random variables U and V such that*

$$\partial_\theta[r_T(Z_T,\theta)] \xrightarrow[T\to\infty]{d} U\sim Q_\theta,$$

and

$$\sqrt{T}(\widehat{\theta}_T-\theta)\xrightarrow[T\to\infty]{d} V\sim R_\theta \quad \textit{with} \quad \widehat{\theta}_T \xrightarrow[T\to\infty]{\mathrm{P}_\theta}\theta.$$

3. *Let $S:\mathbb{R}^+\to\mathbb{R}^+$ be an increasing function such that $\phi(T)-S(T)\to\infty$ and $\frac{S(T)}{T}\to 1$, then*

$$\delta_T = \frac{r_T(Z_T,\widehat{\theta}_{S(T)})-r_T(Z_T,\theta)}{\widehat{\theta}_{S(T)}-\theta} - \partial_\theta r_T(Z_T,\theta) \xrightarrow[T\to\infty]{\mathrm{P}_\theta} 0.$$

In order to get a limit in distribution we first consider a predictor $r_T(Z_T,\widehat{\theta}_{S(T)})$ with an estimator $\widehat{\theta}_{S(T)}$ based on $X_{(S(T))}$.

Proposition 4.1. *Under Assumption 4.1*

$$\sqrt{T}\big(r_T(Z_T,\widehat{\theta}_{S(T)})-r_T(Z_T,\theta)\big)\xrightarrow[T\to\infty]{d} UV$$

with $(U,V)\sim Q_\theta\otimes R_\theta$.

Proof. Let $U_T=\partial_\theta r_T(Z_T,\theta)$ and $V_T=\sqrt{T}(\widehat{\theta}_{S(T)}-\theta)$. The α-mixing condition gives

$$\sup_{\substack{A\in\mathscr{B}_\mathbb{R}\\ B\in\mathscr{B}_\mathbb{R}}} |\mathrm{P}_\theta(U_T\in A, V_T\in B)-\mathrm{P}_\theta(U_T\in A)\mathrm{P}_\theta(V_T\in B)| \leqslant \alpha_\theta\big(\phi(T)-S(T)\big)\xrightarrow[T\to\infty]{} 0$$

Let $A,B \in \mathscr{B}_{\mathbb{R}}$, such that $\mathrm{P}_\theta(U \in \partial A) = 0$ and $\mathrm{P}_\theta(V \in \partial B = 0)$, then

$$|\mathrm{P}_\theta(U_T \in A, V_T \in B) - \mathrm{P}_\theta(U \in A)\mathrm{P}_\theta(V \in B)| \leqslant |\mathrm{P}_\theta(U_T \in A, V_T \in B) - \mathrm{P}_\theta(U_T \in A)\mathrm{P}_\theta(V_T \in B)| \\ + |\mathrm{P}_\theta(U_T \in A)\mathrm{P}_\theta(V_T \in B) - \mathrm{P}_\theta(U \in A)\mathrm{P}_\theta(V \in B)|.$$

Therefore (see [1] theorem 3.1 p. 20)

$$(U_T, V_T) \xrightarrow[T\to\infty]{d} (U,V).$$

Now let

$$\Delta_T = \sqrt{T}\left(r_T(Z_T,\widehat{\theta}_{S(T)}) - r_T(Z_T,\theta)\right) = \left(\partial_\theta r_T(Z_T,\theta) + \delta_T\right)\sqrt{T}\left(\widehat{\theta}_{S(T)} - \theta\right) = (U_T + \delta_T)V_T.$$

Condition 3 of Assumption 4.1 states $\delta_T \xrightarrow[T\to\infty]{\mathrm{P}_\theta} 0$. Then (see [1] p. 25)

$$\left((U_T+\delta_T), V_T\right) \xrightarrow[T\to\infty]{d} (U,V).$$

Hence $\Delta_T \xrightarrow[T\to\infty]{d} UV$. □

Now we can replace $S(T)$ with T in the result under some additional assumptions.

Proposition 4.2. *Under Assumptions 3.3 and 4.1, suppose the following two conditions hold,*

1. $U_T = \partial_\theta r_T(Z_T,\theta)$ *is bounded in distribution,* i.e. *there exists a random variable* $\widetilde{U}$ *such that*

$$\mathrm{P}(|U_T| \geqslant y) \leqslant \mathrm{P}(|\widetilde{U}| \geqslant y), \quad \forall y \geqslant 0,$$

2.

$$\delta_T^* = \frac{r_T(Z_T,\widehat{\theta}_T) - r_T(Z_T,\widehat{\theta}_{S(T)})}{\widehat{\theta}_T - \widehat{\theta}_{S(T)}} - \partial_\theta r_T(Z_T,\theta) \xrightarrow[T\to\infty]{\mathrm{P}_\theta} 0.$$

Then,

$$\sqrt{T}\left(r_T(Z_T,\widehat{\theta}_T) - r_T(Z_T,\theta)\right) \xrightarrow[T\to\infty]{d} UV,$$

with $(U,V) \sim Q_\theta \otimes R_\theta$.

Proof. Let

$$\Delta_T^* = \sqrt{T}\left(r_T(Z_T,\widehat{\theta}_T) - r_T(Z_T,\widehat{\theta}_{S(T)})\right) = \left(U_T + \delta_T^*\right)\sqrt{T}\left(\widehat{\theta}_T - \widehat{\theta}_{S(T)}\right),$$

Let

$$A_T = U_T + \delta_T^* \quad \text{and} \quad B_T = \sqrt{T}\left(\widehat{\theta}_T - \widehat{\theta}_{S(T)}\right).$$

And let $\varepsilon > 0$, then

$$\begin{aligned}
\mathrm{P}_\theta\left(|\Delta_T^*| > \varepsilon\right) &= \mathrm{P}_\theta\left(|A_T B_T| > \varepsilon, |A_T| \geqslant a\right) + \mathrm{P}_\theta\left(|A_T B_T| > \varepsilon, |A_T| < a\right) \\
&\leqslant \mathrm{P}_\theta\left(|A_T| \geqslant a\right) + \mathrm{P}_\theta\left(|B_T| > \frac{\varepsilon}{a}\right) \\
&\leqslant \mathrm{P}_\theta\left(|U_T| \geqslant \frac{a}{2}\right) + \mathrm{P}_\theta\left(|\delta_T^*| \geqslant \frac{a}{2}\right) + \mathrm{P}_\theta\left(|B_T| > \frac{\varepsilon}{a}\right) \\
&\leqslant \mathrm{P}_\theta\left(|\widetilde{U}| \geqslant \frac{a}{2}\right) + \mathrm{P}_\theta\left(|\delta_T^*| \geqslant \frac{a}{2}\right) + \mathrm{P}_\theta\left(|B_T| > \frac{\varepsilon}{a}\right).
\end{aligned}$$

By Assumption 3.3,

$$\mathrm{P}_\theta\left(\sqrt{T}|\widehat{\theta}_T-\widehat{\theta}_{S(T)}|>\varepsilon\right)\leqslant\varepsilon^{-2}\mathrm{E}_\theta\left[T\left(\widehat{\theta}_T-\widehat{\theta}_{S(T)}\right)^2\right]\xrightarrow[T\to\infty]{}0.$$

Hence

$$B_T\xrightarrow[T\to\infty]{\mathrm{P}_\theta}0. \tag{13}$$

Let $\eta>0$, then $\exists a>0$ such that

$$\mathrm{P}_\theta\left(|\widetilde{U}|\geqslant\frac{a}{2}\right)<\frac{\eta}{3},$$

and $\exists T_0>0$ such that $\forall T>T_0$, by condition 2 and (13) respectively,

$$\mathrm{P}_\theta\left(|\delta_T^*|\geqslant\frac{a}{2}\right)<\frac{\eta}{3}\quad\text{and}\quad\mathrm{P}_\theta\left(|B_T|>\frac{\varepsilon}{a}\right)<\frac{\eta}{3}.$$

Hence $\mathrm{P}_\theta\left(|\Delta_T^*|>\varepsilon\right)<\eta$. Finally $\Delta_T^*\xrightarrow{\mathrm{P}_\theta}0$. Therefore

$$\sqrt{T}\left(r_T(Z_T,\widehat{\theta}_T)-r_T(Z_T,\theta)\right)=\Delta_T^*+\Delta_T\xrightarrow[T\to\infty]{d}UV.$$

□

Example 4.1. *Ornstein-Uhlenbeck process*
The process is β-mixing, hence it is α-mixing. Condition 2 of Assumption 4.1 and condition 1 of Proposition 4.2 are fulfilled since

$$U_T=\partial_\theta r_T(Z_T,\theta)=-hX_Te^{-\theta h}\sim\mathscr{N}\left(0,\frac{h^2e^{-2\theta h}}{2\theta}\right),$$

and

$$\sqrt{T}(\widehat{\theta}_T-\theta)\xrightarrow{d}\mathscr{N}(0,2\theta).$$

For the condition 3, we have

$$\begin{aligned}\mathrm{P}_\theta(\delta_T>\varepsilon)&=\mathrm{P}_\theta\left(|X_T|\left|\frac{e^{-\widehat{\theta}_{S(T)}}-e^{-\theta h}}{\widehat{\theta}_{S(T)}-\theta}+he^{-\theta h}\right|>\varepsilon\right)\\&\leqslant\mathrm{P}_\theta(|X_T|\geqslant a)+\mathrm{P}_\theta\left(\left|\frac{e^{-\widehat{\theta}_{S(T)}}-e^{-\theta h}}{\widehat{\theta}_{S(T)}-\theta}+he^{-\theta h}\right|>\frac{\varepsilon}{a}\right)\\&\leqslant\mathrm{P}_\theta(|X_0|\geqslant a)+\mathrm{P}_\theta(|\widehat{\theta}_{S(T)}-\theta|\geqslant\alpha)+\mathrm{P}_\theta\left(\left|\frac{e^{-\widehat{\theta}_{S(T)}}-e^{-\theta h}}{\widehat{\theta}_{S(T)}-\theta}+he^{-\theta h}\right|>\frac{\varepsilon}{a},|\widehat{\theta}_{S(T)}-\theta|<\alpha\right).\end{aligned}$$

Now let $\eta>0$, then $\exists a>0$ such that $\mathrm{P}_\theta(|X_0|\geqslant a)<\frac{\eta}{2})$. There exists $\alpha>0$ such that for all $\theta'\in\mathbb{R}^+$

$$|\theta'-\theta|<\alpha\Rightarrow\left|\frac{e^{-\theta' h}-e^{-\theta h}}{\theta'-\theta}+he^{-\theta h}\right|\leqslant\frac{\varepsilon}{a}.$$

And $\exists T_0>0$, $\forall T>T_0$, $\mathrm{P}_\theta(|\widehat{\theta}_{S(T)}-\theta|\geqslant\alpha)<\frac{\eta}{2}$. Thus $\mathrm{P}_\theta(\delta_T>\varepsilon)<\eta$ ($\widehat{\theta}_T$ is consistent, see [6]). Hence condition 3 of Assumption 4.1 is fulfilled. Proof for condition 2 of Proposition 4.2 is similar.

4.2. *Asymptotic efficiency*

Convergence in distribution of $\partial_\theta r_T(Z_T,\theta)$ only depends on the statistical model. It is therefore natural to define, under conditions of Proposition 4.2, the asymptotic QEP of the predictor $r_T(Z_T,\widehat{\theta}_T)$ as

$$\mathrm{E}_\theta(U^2V^2) = \mathrm{E}_\theta(U^2)\mathrm{E}_\theta(V^2).$$

We see here that it is natural to define that a plug-in predictor is asymptotically efficient if and only if it is based on an asymptotically efficient estimator of θ. Provided that asymptotic efficiency of the estimator of θ is defined with respect to its asymptotic variance (*i.e.* $\mathrm{E}_\theta V^2$) and that appropriate conditions to avoid supereffiency are fulfilled (see for instance [8] section 6.2 or [12] section 8.5).

5. Simulations

In the previous sections it has been seen that, when the assumptions are fulfilled, the asymptotic efficiency of the estimator of the parameter ensures the asymptotic efficiency of the corresponding plug-in predictor. In particular the problem of prediction of the Ornstein-Uhlenbeck process fulfills those assumptions. Hence a plug-in predictor based on an asymptotically efficient estimator like the MLE is an asymptotically efficient predictor. We will call this plug-in predictor the MLP.

The results of the previous sections are limit theorems and they do not tell anything about how large the sample need to be in order to have the QEP get close to the asymptotic bound of efficiency, where we call asymptotic bound of efficiency (ABE) the function

$$T \mapsto \frac{U(\theta)}{TI(\theta)} = \frac{h^2e^{-2\theta h}}{T},$$

where

$$I(\theta) = \lim_{T\to\infty} \frac{I_T(\theta)}{T} = \frac{1}{2\theta} \quad \text{and} \quad U(\theta) = \lim_{T\to\infty} \mathrm{E}_\theta\left(\partial_\theta r_T(X_T,\theta)\right)^2 = \frac{h^2e^{-2\theta h}}{2\theta}.$$

In order to check visually how quickly the QEP of the predictor gets close to the ABE, we performed Monte-Carlo simulations of the Ornstein-Uhlenbeck process with $\theta \in \{0.1, 0.3, 1, 3\}$ and computed the error of its (approximated) MLP for $h \in \{1,3\}$. We also compute the QER with respect to the distribution of X_T.

A program was written by the authors in the OCaml language to carry out the calculations of the simulated sample paths, the predictor, its QEP and QER and the ABE. The number of simulated paths of the process for each value of θ is 10^6. Paths of the Ornstein-Uhlenbeck process were generated with a discretization step of 0.05. The discretized sample paths were computed from the innovation processes, which are i.i.d. sequences of gaussian samples. They were generated using the `zigurrat` algorithm of the GNU Scientific Library (GSL) with the pseudo-random number generator `taus` of the same library. The graphical output was performed with the R programming language.

The MLE of θ is

$$\widehat{\theta}_T = \frac{T - X_0^2 - X_T^2 + \sqrt{(T - X_0^2 - X_T^2)^2 + 8\int_0^T X_t^2 dt}}{4\int_0^T X_t^2 dt},$$

and we use the following discretized approximation instead

$$\widetilde{\theta}_T = \frac{T - X_0^2 - X_T^2 + \sqrt{(T - X_0^2 - X_T^2)^2 + 8s\sum_{n=0}^{T/s} X_{sn}^2}}{4s\sum_{n=0}^{T/s} X_{sn}^2},$$

with s, the size of the discretization step, is such that $s = \frac{1}{r} = 0.05$ with the sampling rate $r = 20$. Off course the MLE is not computable since we always get a finite number of pieces of data and never a continuum of observations.

Figures 1 and 2 show plots of the QEP and the QER of the MLP in thin black and blue lines respectively, plotted against the time T at which the prediction of $\mathrm{E}_\theta[X_{T+h}|X_T]$ is performed, along with the ABE in thick red line. Logarithmic scales are used for both axis of the plots to make them easier to interpret.

The QEP stands less favorably with respect to the ABE as θ and h get large. The QER does not show this pattern. For all the values of h and θ tested, the QEP and the QER finally attain a regime of asymptotic efficiency at some point. The time at which the QEP, the QER and the ABE become of the same order of magnitude varies according to the parameter θ and the horizon of prediction h. The QEP attains this regime more quickly for $h = 1$ than $h = 3$. But otherwise, no clear relationship appears between the time at which the asymptotic regime is attained and the values of the parameters.

We also note that the QEP and the QER converge to the ABE even though the estimator we used, in the plug-in predictor, is only a discretized approximation of the MLE.

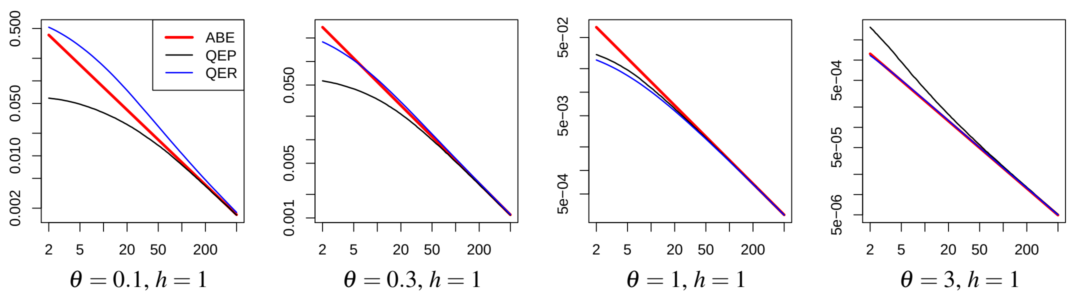

FIGURE 1. *ABE, QEP and QER plotted against time, for $h = 1$*

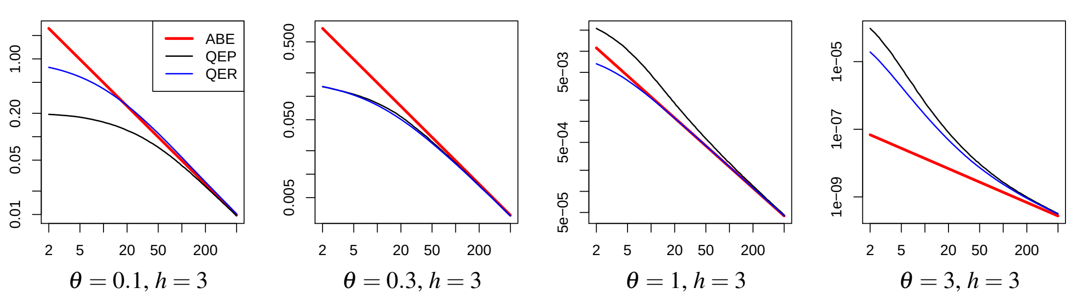

FIGURE 2. *ABE, QEP and QER plotted against time, for $h = 3$*

Acknowledgement

We thank the anonymous referee for the thorough and detailed report on the first version of the manuscript which lead to substantial improvements of the article.

References

[1] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons Inc., New York, 1968.
[2] Denis Bosq. Exact asymptotic bias for estimators of the Ornstein-Uhlenbeck process. *Stat. Inference Stoch. Process.*, 13(2):133–145, 2010.
[3] Denis Bosq and Delphine Blanke. *Inference and prediction in large dimensions*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester, 2007.
[4] Paul Doukhan. *Mixing: Properties and examples*, volume 85 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1994.
[5] I. A. Ibragimov and R. Z. Has′minskiĭ. *Statistical estimation*, volume 16 of *Applications of Mathematics*. Springer-Verlag, New York, 1981. Asymptotic theory, Translated from the Russian by Samuel Kotz.
[6] Yury A. Kutoyants. *Statistical inference for ergodic diffusion processes*. Springer Series in Statistics. Springer-Verlag London Ltd., London, 2004.
[7] Lucien Le Cam. *Asymptotic methods in statistical decision theory*. Springer Series in Statistics. Springer-Verlag, New York, 1986.
[8] E. L. Lehmann and George Casella. *Theory of point estimation*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 1998.
[9] Tapan K. Nayak. On best unbiased prediction and its relationships to unbiased estimation. *J. Statist. Plann. Inference*, 84(1-2):171–189, 2000.
[10] Tapan K. Nayak. Rao-Cramer type inequalities for mean squared error of prediction. *Amer. Statist.*, 56(2):102–106, 2002.
[11] Emmanuel Onzon. Multivariate Cramér-Rao inequality for prediction and efficient predictors. *Statistics & Probability Letters*, 81(3):429–437, 2011.
[12] A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.
[13] A. Yu. Veretennikov. Bounds for the mixing rate in the theory of stochastic equations. *Theory of Probability and its Applications*, 32(2):273–281, 1987.
[14] Yannis G. Yatracos. On prediction and mean squared error. *Canad. J. Statist.*, 20(2):187–200, 1992.