ASYMPTOTICALLY EFFICIENT PREDICTION
FOR LAN FAMILIES

BY EMMANUEL ONZON,
Université Lyon 1

Abstract In a previous paper (Bosq and Onzon (2012)) we did a first generalization of the concept of asymptotic efficiency for statistical prediction, i.e. for the problems where the unknown quantity to infer is not deterministic but random. However, in some instances, the assumptions we made were not easy to verify. Here we give proofs of similar results based on quite a different set of assumptions. The model is required to be a LAN family, which allows to use the convolution theorem of Hájek and Le Cam. The results are applied to the forecasting of a bivariate Ornstein-Uhlenbeck process, for which the assumptions of (Bosq and Onzon (2012)) are tricky to verify.

1. Introduction.

1.1. Background. The theory of statistical prediction is an extension of the theory of point estimation where the unknown quantity to infer from the observation is not deterministic but random. It develops by generalizing the results of point estimation to prediction problems (see for instance Yatracos (1992), Bosq and Blanke (2007) and Bosq (2007)). For instance some authors have studied the extension of the Cramér-Rao inequality to the case of statistical prediction problems (Yatracos (1992), Miyata (2001), Nayak (2002), Onzon (2011)). In the univariate case and for an unbiased predictor $p(X)$ it has the form

\[(1.1) \quad E_\theta(p(X) - r(X, \theta))^2 \geq \frac{(E_\theta(\partial_\theta r(X, \theta)))^2}{I(\theta)},\]

where $I(\theta)$ is the Fisher information pertaining to the observation $X$, and $r(X, \theta)$ is the conditional expectation of the unobserved variable $Y$ to predict given the observation $X$,

\[r(X, \theta) = E_\theta[Y|X].\]
The Cramér-Rao inequality gives a lower bound on the mean squared error of estimation which gives an optimality criterion with which to compare the performance of a given estimator. One limitations of this lower bound, however, is the fact that it depends on the bias of the estimator which is applied to. This means that the optimality criterion it provides only allows to compare estimators with the same bias. For this reason, the Cramér-Rao inequality is mainly useful for unbiased estimation. Another limitation is the fact that it is not necessarily attained by any estimator at all, and an estimator which risk does not attains the bound might still be the optimal one.

The approach that was developed to overcome these limitations, consists in studying the limit of the (normalized) error of estimation when the sample size tends to infinity. This work culminated with the celebrated Hájek–Le Cam convolution theorem and the Hájek–Le Cam minimax inequality (Hájek (1969/1970), Hájek (1972), Le Cam (1972)).

The two limitations of the Cramér-Rao inequality discussed above also hold for the extension of the inequality to prediction. Moreover in some instances the lower bound given by the inequality for predictors is zero, which does not give any information at all. This is the case for example in the problem of forecasting the Ornstein-Uhlenbeck process, or more generally, as soon as $E_{\theta}(\partial_{\theta} r(X, \theta)) = 0$. It has been seen in Bosq and Onzon (2012) that the asymptotic approach pursued for estimation could be generalized to prediction problems under some suitable conditions. Besides, the asymptotic lower bound obtained is of the form

$$\frac{E_{\theta}(\partial_{\theta} r(X, \theta))^2}{I(\theta)},$$

instead of the form of (1.1). The square being inside the expectation, the lower bound for problems with $E_{\theta}(\partial_{\theta} r(X, \theta)) = 0$ is not zero anymore. In the case of the problem of forecasting the univariate stationary Ornstein-Uhlenbeck process, Bosq and Onzon (2012) prove that the asymptotic lower bound is attained by the plug-in predictor with the parameter $\theta$ being estimated by the maximum likelihood estimator (MLE). The results in Bosq and Onzon (2012) are based on a version of the Cramér-Rao bound for the estimation of the regression function $x \mapsto r(x, \theta)$. This lower bound depends on the bias $b(x, \theta)$ of estimation of the regression function and its derivatives $\partial_{\theta} b(x, \theta)$. A consequence is that the asymptotic results require the latter to converge towards zero fast enough. In the case of the forecasting of the univariate Ornstein-Uhlenbeck process, this condition is verified for the plug-in estimator $r(\cdot, \hat{\theta})$ where $\hat{\theta}$ is the MLE. This follows from results in
Bosq (2010) that are derived from the closed form of the MLE \( \hat{\theta} \). However, depending on the problems, the condition on the speed of convergence of the derivatives of the bias is not always easy to verify. For instance, in the case of the bivariate Ornstein-Uhlenbeck process, no closed form is known for the MLE. Hence the results of Bosq (2010) cannot be generalized directly to the multivariate case. Consequently the verification of the condition on the speed of convergence of the derivatives of the bias is not easy to verify. In this article we present results that rely of a different set of assumptions which are fulfilled in the case of the forecasting of the bivariate Ornstein-Uhlenbeck process. The asymptotic lower bounds presented here are deduced from the Hájek–Le Cam convolution theorem.

In the remainder of this section we set the notations that will be useful throughout the article, pose the problem we are interested in and then give a summary of the main results presented here.

1.2. Notations. We will use the following notations. For all \( m \) and \( n \) positive integers let \( M_{m,n} \) be the space of matrices with \( m \) lines and \( n \) columns with real coefficients. We will use the Frobenius norm on this space, which we will denote \( \| \cdot \|_{M_{m,n}} \), and is defined as follows. If \( A \in M_{m,n} \) with coefficients \( a_{i,j}, 1 \leq i \leq m, 1 \leq j \leq n \) then

\[
\| A \|_{M_{m,n}} = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{i,j}^2} = \sqrt{\text{trace}(A^t A)} = \sqrt{\sum_{i=1}^{n} \sigma_i^2}
\]

where \( \sigma_i^2 \) are the eigenvalues of \( A^t A \) and where \( A^t \) is the transpose matrix of \( A \). For square matrices, \( M_n = M_{n,n} \). For a matrix \( A \in M_{m,n} \), we will note \( AA^t = A^2 \).

The symbols \( J_\theta \) and \( \nabla_\theta \) represent respectively, the jacobian matrix and gradient operators, with respect to the multidimensional variable \( \theta \in \mathbb{R}^d \). The symbol \( \partial_\theta \) represents the differentiation with respect to \( \theta \) when \( \theta \) is a real variable.

When we write an inequality between matrices, it will always refer to Löwner partial order, i.e. \( A \preceq B \) if and only if \( A \) and \( B \) are square matrices of same size and \( B - A \) is a symmetric positive semidefinite matrix.

1.3. Model and problem. Let \( (X_t, t \in \mathcal{T}) \), where \( \mathcal{T} = \mathbb{N} \) or \( \mathcal{T} = \mathbb{R}_+ \), be a random process taking its values in \( E \) where \( (E, \mathcal{B}) \) is a measurable space. We will note \( (X_t, t \geq 0) \) without refering to \( \mathcal{T} \) thereafter. We denote its distribution by \( P_\theta \) where \( \theta \in \Theta \subset \mathbb{R}^d \) is an unknown parameter. For all time \( T \geq 0 \), we assume that the path

\[
X_{(T)} = (X_t, 0 \leq t \leq T),
\]
is observed. Let \((Y_t, t \geq 0)\) be a random process taking its values in \(\mathbb{R}^k\), such that for all \(T \geq 0\), \(Y_T\) is not observed at time \(T\). In the application we have in mind, \(Y_T = X_{T+h}\) with \(h > 0\).

In this section we consider the problem of estimating the following regression function,
\[
r(\cdot, \theta) : x \mapsto E_\theta(Y_T | X_T = x),
\]
given \(X(T)\). Here the function \(r\) is assumed not to depend on \(T\), (we did not make this restriction in Bosq and Onzon (2012)). We will consistently assume that for all \(T \geq 0\) the function \((x, \theta) \mapsto r(x, \theta)\) is measurable and for all \(x \in E\), the function \(\theta \mapsto r(x, \theta)\) is differentiable over \(\Theta\).

In the application, we have in mind that this regression function is
\[
r(\cdot, \theta) : x \mapsto E_\theta(X_{T+h} | X_T = x).
\]

The risk we choose for this problem is the quadratic error of estimation of the regression function (QER),
\[
\rho_T(\theta) = \int_E E_\theta(\hat{r}_T(x) - r(x, \theta))^2 d\mu_\theta(x),
\]
where we choose \(\mu_\theta\), a measure over \((E, \mathcal{B})\), with \(E\) the range of the random variable \(X_T\), and \(\hat{r}_T\) an estimator of the regression function \(r(\cdot, \theta)\). Oftentimes \(\mu_\theta\) will be the distribution of \(X_T\) under \(P_\theta\), and \(\hat{r}_T\) will be a plug-in estimator, i.e. of the form \(\hat{r}_T = r(\cdot, \hat{\theta}_T)\). Here the measure \(\mu_\theta\) is assumed not to depend on \(T\) (we did not make this restriction in Bosq and Onzon (2012)).

We also consider the corresponding prediction problem, which consists in predicting
\[
r(X_T, \theta) = E_\theta(Y_T | X_T),
\]
given \(X(T)\). In the application we have in mind this conditional expectation will be
\[
r(X_T, \theta) = E_\theta(X_{T+h} | X_T).
\]
The risk we choose for this problem is the quadratic error of prediction (QEP),
\[
R_T(\theta) = E_\theta(\hat{r}_T(X_T) - r(X_T, \theta))^2.
\]

This prediction problem identifies with prediction of \(Y_T\) when the following condition is met,
\[
E_\theta(Y_T | X(T)) = E_\theta(Y_T | X_T).
\]
For instance, this condition is satisfied for the forecasting problem of a future value $X_{T+h}$ of a Markov process $(X_t, t \geq 0)$ observed until time $T$, taking $Y_T = X_{T+h}$.

**Remark 1.1.** When the relation (1.2) is not satisfied for the problem at hand, it might however be the case that after a recasting of the problem, it is satisfied. For example, when forecasting an AR($p$) process, the conditional expectation is

$$ E_\theta (X_{T+h} | X_{(T)}) = E_\theta (X_{T+h} | X_T, X_{T-1}, \ldots, X_{T-p+1}) \neq E_\theta (X_{T+h} | X_T). $$

We can however set $Z_t = (X_t, \ldots, X_{T-p+1})$ and then

$$ E_\theta (Z_{T+h} | Z_{(T)}) = E_\theta (Z_{T+h} | Z_T). $$

**Remark 1.2.** Throughout the paper we will implicitly assume that the following conditions are fulfilled. For all $x \in E$, $\theta \in \Theta$, $T > 0$,

1. $J_\theta r(x, \theta)$ exists,
2. $r(\cdot, \theta) \in L^2(\mu_\theta)$ and $J_\theta r(\cdot, \theta) \in L^2(\mu_\theta)$.

1.4. **Summary of the results.** In Section 2 we establish an asymptotic lower bound for the QER, for all $y \in \mathbb{R}^k$,

$$ \lim_{T \to \infty} T y' \rho_T(\theta) y \geq y' \int_E (J_\theta r(x, \theta)) I(\theta)^{-1} (J_\theta r(x, \theta))' d\mu_\theta(x) y, $$

and deduce a definition of asymptotic efficiency for the corresponding problem of estimation of the regression function.

Under the assumption of asymptotic equivalence between the risks of both problems,

$$ (1.3) \quad \lim_{T \to \infty} T \rho_T(\theta) = \lim_{T \to \infty} T R_T(\theta), $$

an asymptotic lower bound on the QEP immediately ensues, for all $y \in \mathbb{R}^k$,

$$ \lim_{T \to \infty} T y' R_T(\theta) y \geq y' \int_E (J_\theta r(x, \theta)) I(\theta)^{-1} (J_\theta r(x, \theta))' d\mu_\theta(x) y. $$

From this bound we deduce a definition of asymptotic efficiency for the prediction problem.

In Section 3 we give conditions allowing to verify the asymptotic equivalence (1.3) in the case of plug-in predictors, i.e. of the form $r(X_T, \hat{\theta}_T)$, where
\( \hat{\theta}_T \) is an estimator of \( \theta \) based on \( X(T) \). Theorems 3.1 and 3.2, proved in Section 3.2, state conditions that imply (1.3). The conditions of Theorem 3.2 are fulfilled in our example of forecasting the bivariate Ornstein-Uhlenbeck process.

Finally in Section 3.3 we see that under some conditions the limit of the QER can be written, in the univariate and multivariate cases respectively,

\[
\lim_{T \to \infty} T \rho_T(\theta) = \int_E \left( \partial_\theta r(x, \theta) \right)^2 d\mu_\theta(x) V(\theta), \\
\lim_{T \to \infty} T \rho_T(\theta) = \int_E \left( J_\theta r(x, \theta) \right) V(\theta) \left( J_\theta r(x, \theta) \right)' d\mu_\theta(x),
\]

with \( V(\theta) = \lim_{T \to \infty} T E_\theta(\hat{\theta}_T - \theta)^2 \). The result implies the fulfillment of Assumption 3.3 which is required by Theorems 3.1 and 3.2.

A remarkable consequence of this result is that, under its conditions, the asymptotic efficiency of a plug-in predictor \( r(X_T, \hat{\theta}_T) \) comes down to the asymptotic efficiency of the estimator \( \hat{\theta}_T \).

2. Asymptotic efficiency for regression and prediction. We begin by reminding the convolution theorem of Hájek and Le Cam and other propositions that will be useful in proving the results of the article.

2.1. LAN families and Hájek–Le Cam convolution theorem. In this section we gather definitions and results pertaining to locally asymptotically normal (LAN) families and asymptotic efficiency for reference in the remainder of the article.

References about the theory of asymptotic efficiency are the books of Ibragimov and Has’minski˘ı (1981), Le Cam (1986), Pfanzagl (1994), Bickel et al. (1998) and van der Vaart (1998).

Definition 2.1. Let \( \Theta \subset \mathbb{R}^d \) be the domain of the parameter \( \theta \). A family of probability measures \( \left( P^T_\theta, \theta \in \Theta \right) \), indexed by \( T \in \mathbb{R}_+ \) or \( T \in \mathbb{N} \), is said locally asymptotically normal (LAN) at \( \theta \in \Theta \), if there is a family of positive real numbers \( (c_T, T \geq 0) \) such that \( c_T \to 0 \) as \( T \to \infty \), a symmetric positive definite matrix \( I(\theta) \) and a family of random vectors \( (\Delta_T(\theta), T \geq 0) \), such that for all \( u \in \mathbb{R}^d \),

\[
\log \frac{dP^T_{\theta+c_Tu}}{dP^T_\theta}(X(T)) = u' \Delta_T(\theta) - \frac{1}{2} u' I(\theta) u + \varepsilon_T(X(T), \theta, u),
\]

and the following convergences hold.
ASYMPTOTICALLY EFFICIENT PREDICTION

1. \( P_\theta \circ \Delta_T(\theta) \xrightarrow{T \to \infty} \mathcal{N}(0, I(\theta)) \)

2. \( P_\theta - \lim_{T \to \infty} \varepsilon_T(X_T(\theta), \theta, u) = 0. \)

In the previous definition we used the following notations. For all probability measures \( P \) and all random variables \( X, P \circ X \) represents the distribution of \( X \) under \( P \). The notation \( \Rightarrow_{T \to \infty} \) is used for weak convergence (convergence in distribution).

\textbf{Remark 2.1.} Numerous families of random processes are LAN. In particular, the property holds for ergodic diffusions (Kutoyants (2004)), jump Markov processes (Höpfner (1988), Höpfner, Jacod and Ladelli (1990)), and also for ARMA models (Akritas and Johnson (1982), Swensen (1985), Kreiss (1987), Garel and Hallin (1995)), and many other time series models (for instances Koul and Schick (1997), Taniguchi and Kakizawa (2000), Drost, van den Akker and Werker (2008)).

\textbf{Definition 2.2.} Assume \((P^T_\theta, \theta \in \Theta), T \geq 0\) is a LAN family. Let \( \hat{\psi}_T, T \geq 0 \) be a family of estimators of \( \psi(\theta) \) such that for all \( T \geq 0 \), \( \hat{\psi}_T \) is \( X_T(\cdot) \)-measurable. We say \( \hat{\psi}_T, T \geq 0 \) is a family of estimators regular at \( \theta \in \Theta \), if there is a distribution \( L(\theta) \) such that for all \( u \in \mathbb{R}^d \),

\[
P^T_{\theta + c_T u} \circ \Delta_T^{-1} \left( \hat{\psi}_T - \psi(\theta + c_T u) \right) \xrightarrow{T \to \infty} L(\theta).
\]

We state the convolution theorem proved independently by Hájek and Le Cam (Hájek (1969/1970), Le Cam (1972)). This result will be useful in Section 2. We give here the statement of Pfanzagl (1994).

\textbf{Theorem 2.1.} Let \((P^T_\theta, \theta \in \Theta), T \geq 0, \) be a LAN family, with \( \Theta \subseteq \mathbb{R}^d \). Let \( \psi : \Theta \to \mathbb{R}^k, k \leq d, \) be a differentiable function with Jacobian matrix \( J_\theta \psi(\theta) \) with rank \( k \) for all \( \theta \in \Theta \). Let \( \hat{\psi}_T, T \geq 0, \) be a family of estimators regular at every \( \theta \in \Theta \). Then for all \( \theta \in \Theta \) there is a distribution \( M(\theta) \) over \((\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))\) such that

\[
(2.1) \quad P^T_\theta \circ \left( \Delta_T(\theta), c_T^{-1}(\hat{\psi}_T - \psi(\theta)) - (J_\theta \psi(\theta))I(\theta)^{-1}\Delta_T(\theta) \right) \xrightarrow{T \to \infty} \mathcal{N}(0, I(\theta)) \times M(\theta).
\]

This implies

\[
L(\theta) = \mathcal{N}(0, (J_\theta \psi(\theta))I(\theta)^{-1}(J_\theta \psi(\theta))') \ast M(\theta).
\]
The following proposition will allow us to show that a family of estimators is regular. It is a consequence of Proposition 7.1.8 p.227 of Pfanzagl (1994).

**Proposition 2.1.** Let \( \Theta \subset \mathbb{R}^d \) and \( (P_T^\theta, \theta \in \Theta) \), be a LAN family indexed by \( T \geq 0 \). Let \( \psi : \Theta \to \mathbb{R}^k \) and \( \hat{\psi}_T, T \geq 0 \), be a family of estimators. Assume there is a family of distributions \( (L(\theta), \theta \in \Theta) \) such that the following conditions are fulfilled.

1. For all \( \theta \in \Theta \), there is a neighbourhood \( V(\theta) \) of \( \theta \) such that the following convergence holds uniformly over \( V(\theta) \).
   \[
P_T^\theta \circ c_T^{-1}\left( \hat{\psi}_T - \psi(\cdot) \right) \xrightarrow{T \to \infty} L(\theta)
\]

2. \( \theta \mapsto L(\theta) \) is continuous for weak convergence, i.e., for all \( \theta_0 \in \Theta \),
   \[
   L(\theta) \xrightarrow{\theta \to \theta_0} L(\theta_0).
   \]

Then the family of estimators \( \hat{\psi}_T, T \geq 0 \), is regular at every \( \theta \in \Theta \).

In the previous proposition we have used the notion of uniform weak convergence that is defined as follows.

**Definition 2.3.** Let \( \Theta \subset \mathbb{R}^d \) and \( Q^{(n)}_\theta \), for all \( n \in \mathbb{N}, \theta \in \Theta \), let \( Q^{(n)}_\theta \) and \( Q_\theta \) be probability measures over \((\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))\). The sequence \( Q^{(n)}_\theta, n \in \mathbb{N} \) converges weakly to \( Q_\theta \) uniformly over \( \Theta \) if for any bounded continuous function \( h \),
   \[
   \lim_{n \to \infty} \sup_{\theta \in \Theta} \left| Q^{(n)}_\theta(h) - Q_\theta(h) \right| = 0.
   \]

The following proposition (Pfanzagl (1994) p.289) gives an asymptotic expansion for a family of estimators which has the optimal limit in distribution.

**Proposition 2.2.** Under conditions of Theorem 2.1, any regular family of estimators \( (\hat{\psi}_T, T \geq 0) \) which has the optimal limit in distribution, i.e.
   \[
   L(\theta) = \mathcal{N}(0, (J_\theta \psi(\theta)) I(\theta)^{-1} (J_\theta \psi(\theta))^\prime),
   \]
has the following asymptotic expansion. As \( T \to \infty \),
   \[
   \sqrt{T}(\hat{\psi}_T - \psi(\theta)) = (J_\theta \psi(\theta)) I(\theta)^{-1} \Delta_T(\theta) + o_{P_T^\theta}(1).
   \]
2.2. Asymptotic bounds for LAN families. Assumptions 2.1 and 2.2 we made in the article (Bosq and Onzon (2012)), in particular about the asymptotic behaviour of the derivative of the bias, may seem somewhat arbitrary and may be difficult to verify in some instances. In this article we present an alternative approach which allows to dispense with those by using other assumptions. We begin with the following assumption.

**Assumption 2.1.**

1. The family of models $(P_T^0, \theta \in \Theta), T \geq 0$, where $P_T^0$ is the distribution of $(X_t, 0 \leq t \leq T)$ under $P_\theta$, is LAN at every $\theta \in \Theta$, with $c_T = T^{-1/2}$.

2. For all $\theta \in \Theta$, for $\mu_\theta$-almost all $x \in E$, the family of estimators $\hat{r}_T(x)$ is regular at $\theta$ for estimating $r(x, \theta)$ and the limit in distribution is centered with finite variance, i.e. there is a centered distribution $L(x, \theta)$ with finite variance such that the following convergence holds for all $u \in \mathbb{R}^d$

$$P_T^{\theta + u/\sqrt{T}} \circ \sqrt{T} \left( \hat{r}_T(x) - r \left( x, \theta + u/\sqrt{T} \right) \right) \Rightarrow L(x, \theta).$$

Condition (2.2), which corresponds to regularity of the family of estimators, makes possible to apply the Theorem 2.1 (Hájek–Le Cam convolution theorem).

**Lemma 2.1.** Let $Y_n, n \in \mathbb{N}$ be a sequence of random variables of $\mathbb{R}$ such that $Y_n \Rightarrow Y$ where $EY = 0$ and $EY^2 = v$. If $EY_n^2 \overset{n \to \infty}{\longrightarrow} w$ then $w \geq v$.

**Proof.** This result follows from Lemma 1.14 p. 437 from Lehmann and Casella (1998).

**Proposition 2.3.** In the univariate case and under Assumption 2.1, the following asymptotic lower bound holds for all $\theta \in \Theta$,

$$\lim_{T \to \infty} T \rho_T(\theta) \geq \int_E \frac{\left( \partial \theta r(x, \theta) \right)^2}{I(\theta)} d\mu_\theta(x).$$

**Proof.** Let $x \in E$ such that condition 2 of Assumption 2.1 is fulfilled. According to Hájek–Le Cam convolution theorem, for all $\theta \in \Theta$, there is $M_x(\theta)$ such that

$$L(x, \theta) = \mathcal{N} \left( 0, \frac{\left( \partial \theta r(x, \theta) \right)^2}{I(\theta)} \right) * M_x(\theta).$$
Using Lemma 2.1 we deduce
\[ \lim_{T \to \infty} T E_\theta(\hat{r}_T(x) - r(x, \theta))^2 \geq \frac{\left( \partial_\theta r(x, \theta) \right)^2}{I(\theta)}. \]

Now, from Fatou lemma,
\[ \lim_{T \to \infty} \int_E T E_\theta(\hat{r}_T(x) - r(x, \theta))^2 d\mu_\theta(x) \geq \int_E \lim_{T \to \infty} T E_\theta(\hat{r}_T(x) - r(x, \theta))^2 d\mu_\theta(x), \]
therefore
\[ \lim_{T \to \infty} T \rho_T(\theta) \geq \int_E \frac{\left( \partial_\theta r(x, \theta) \right)^2}{I(\theta)} d\mu_\theta(x). \]

**Remark 2.2.** It is desirable that the quantities \( \lim_{T \to \infty} T \rho_T(\theta) \) and \( \lim_{T \to \infty} T \rho_T(\theta) \) be as small as possible. Yet, from Proposition 2.3, they are at least equal to
\[ \int_E \frac{\left( \partial_\theta r(x, \theta) \right)^2}{I(\theta)} d\mu_\theta(x). \]
Hence Proposition 2.3 gives an asymptotic optimality criterion for the estimators of the regression function in the univariate case and in the setting of Assumption 2.1. The following definition ensues.

**Definition 2.4.** We say that a family of estimators \( \hat{r}_T, T \geq 0 \), is asymptotically efficient for estimating \( r(\cdot, \theta) \) if for all \( \theta \in \Theta \)
\[ \lim_{T \to \infty} T \rho_T(\theta) = \int_E \frac{\left( \partial_\theta r(x, \theta) \right)^2}{I(\theta)} d\mu_\theta(x). \]

As a shortcut, we will say that the estimator \( \hat{r}_T \) is asymptotically efficient, to mean that the family of estimators \( \hat{r}_T, T \geq 0 \), is asymptotically efficient.

**Assumption 2.2.** We assume that the following limits exist, are finite, non zero and equal.
\[ \lim_{T \to \infty} T \rho_T(\theta) = \lim_{T \to \infty} T R_T(\theta). \]

For the application that we have in mind, forecasting of random processes, it is not straightforward to verify this assumption. Theorems 3.1 and 3.2, derived in Section 3.2, give conditions to verify this assumption. Theorem 3.2 will allow us to verify Assumption 2.2 for the problem of forecasting of the bivariate stationary Ornstein-Uhlenbeck process in Section 3.2.3.

The following corollary is deduced immediately.
**Corollary 2.1.** In the univariate case, under Assumptions 2.1 and 2.2, for all \( \theta \in \Theta \)

\[
\lim_{T \to \infty} TR_T(\theta) \geq \int_E \frac{(\partial \theta r(x, \theta))^2}{I(\theta)} d\mu_\theta(x).
\]

This result provides an asymptotic optimality criterion for the predictors of \( r(X_T, \theta) \) in the univariate case and in the setting of Assumptions 2.1 and 2.2. The following definition ensues.

**Definition 2.5.** We say that a family of predictors \( \hat{r}(X_T), T \geq 0 \), is asymptotically efficient for predicting \( r(X_T, \theta) \) if for all \( \theta \in \Theta \),

\[
\lim_{T \to \infty} TR_T(\theta) = \int_E \frac{(\partial \theta r(x, \theta))^2}{I(\theta)} d\mu_\theta(x).
\]

We now continue with the generalization of these results to the multivariate case.

**Lemma 2.2.** Let \( Y_n, n \in \mathbb{N} \) be a sequence of random vectors of \( \mathbb{R}^k \) such that \( Y_n \Rightarrow Y \) where \( EY = 0 \) and \( EY \times EY = V \). Let \( y \in \mathbb{R}^k \), if \( y'(EY_n \times EY_n)y \xrightarrow{n \to \infty} w \) then \( w \geq y'Vy \).

**Proof.** The result follows from Lemma 2.1 applied to the sequence of real random variables \( (y'Y_n), n \in \mathbb{N} \). \( \square \)

**Proposition 2.4.** Under Assumption 2.1, the following asymptotic lower bound holds for all \( \theta \in \Theta \) and all \( y \in \mathbb{R}^k \)

\[
\lim_{T \to \infty} Ty'\rho_T(\theta)y \geq y' \int_E (J_\theta r(x, \theta))I(\theta)^{-1}(J_\theta r(x, \theta))' d\mu_\theta(x)y.
\]

**Proof.** The proof is identical to the proof of Proposition 2.3 except we apply Lemma 2.2 instead of Lemma 2.1. \( \square \)

**Remark 2.3.** The smaller \( \lim_{T \to \infty} Ty'\rho_T(\theta)y \) and \( \lim_{T \to \infty} Ty'\rho_T(\theta)y \) are, the better the estimator is (asymptotically). Yet, from Proposition 2.4, they are at least equal to

\[
y' \int_E (J_\theta r(x, \theta))I(\theta)^{-1}(J_\theta r(x, \theta))' d\mu_\theta(x)y.
\]

Hence Proposition 2.4 gives an asymptotic optimality criterion for the estimators of the regression function in the multivariate case and in the setting of Assumption 2.1. The following definition ensues.
Definition 2.6. We say that the family of estimators $\hat{r}_T$, $T \geq 0$, is asymptotically efficient for estimating $r(\cdot, \theta)$ if for all $\theta \in \Theta$,

$$
\lim_{T \to \infty} T \rho_T(\theta) = \int_E (J_{\theta r}(x, \theta)) I(\theta)^{-1}(J_{\theta r}(x, \theta))' d\mu_\theta(x).
$$

In the multivariate case as well as in the univariate case, Assumption 2.2 allows to obtain immediately an asymptotic bound on QEP.

Corollary 2.2. In the multivariate case, under Assumptions 2.1 and 2.2, for all $\theta \in \Theta$

$$
\lim_{T \to \infty} Ty' R_T(\theta)y \geq y' \int_E (J_{\theta r}(x, \theta)) I(\theta)^{-1}(J_{\theta r}(x, \theta))' d\mu_\theta(x)y.
$$

The following generalization of Definition 2.5 to the multivariate case ensues.

Definition 2.7. We say that a family of predictors $\hat{r}(X_T)$, $T \geq 0$, is asymptotically efficient for predicting $r(X_T, \theta)$ if for all $\theta \in \Theta$,

$$
\lim_{T \to \infty} TR_T(\theta) = \int_E (J_{\theta r}(x, \theta)) I(\theta)^{-1}(J_{\theta r}(x, \theta))' d\mu_\theta(x).
$$

2.3. Bivariate Ornstein-Uhlenbeck process. Here we study a forecasting problem of a bivariate stationary Ornstein-Uhlenbeck process. The regression function is estimated by a plug-in estimator for which the model parameter is estimated by the MLE (maximum likelihood estimator). We are going to see that Assumption 2.1 is fulfilled and the plug-in estimator is asymptotically efficient.

2.3.1. Model. We consider the stationary solution $(X_t \in \mathbb{R}^2$, $t \geq 0)$ of the following stochastic differential equation,

$$
dX_t = -Q(\theta)X_t dt + dW_t,
$$

with $(W_t, t \geq 0)$ a standard bivariate Wiener process (i.e. a bivariate random process which components are two independent Wiener processes) and,

$$
Q(\theta) = \alpha I_2 + \beta A = \begin{pmatrix} \alpha \\ \beta \\ \alpha \end{pmatrix} \quad \text{with} \quad A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \theta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}.
$$

We fix the following set of parameters,

$$
\Theta = \{ \theta \in \mathbb{R}^2 \mid Q(\theta) = \alpha I_2 + \beta A \text{ is positive definite} \}.
$$
Applying Itô formula to $g(t, x) = e^{Q(\theta)t}x$ gives an expression of $X_T$,

$$X_T = e^{-Q(\theta)^T}X_0 + \int_0^T e^{Q(\theta)(t-T)}dW_t.$$  

We derive the variance of $X_T$.

$$E_\theta X_T^2 = e^{-Q(\theta)^T}(E_\theta X_0^2)e^{-Q(\theta)^T} + E_\theta \left( \int_0^T e^{-Q(\theta)(t-T)}dW_t \right)^2$$

$$= e^{-Q(\theta)^T}(E_\theta X_0^2)e^{-Q(\theta)^T} + \int_0^T e^{2Q(\theta)(t-T)}dt$$

$$= e^{-Q(\theta)^T}(E_\theta X_0^2)e^{-Q(\theta)^T} + \frac{Q(\theta)^{-1}}{2} - \frac{Q(\theta)^{-1}}{2} e^{-2Q(\theta)T}.$$  

For a stationary process it follows that

$$E_\theta X_T^2 = \frac{Q(\theta)^{-1}}{2} = \frac{1}{2(\alpha^2 - \beta^2)} \begin{pmatrix} \alpha & -\beta \\ -\beta & \alpha \end{pmatrix}.$$  

For any $\theta \in \Theta$ the distribution of $X(T)$ is absolutely continuous with respect to the distribution of the process $(U + W_t, t \geq 0)$ where $U$ is a standard gaussian random vector independent from $(W_t, t \geq 0)$, let $\nu$ be its distribution. The density of $X(T)$ with respect to $\nu$ is the function $f_T(\cdot, \theta) : \mathcal{X}_T \to \mathbb{R}$

$$f_T(X(T), \theta) = 2\sqrt{\det Q(\theta)} \times$$

$$\exp \left\{ \alpha T - \frac{1}{2}(X'_0Q(\theta)X_0 + X'_TQ(\theta)X_T) - \frac{1}{2} \int_0^T \|Q(\theta)X_t\|^2 dt \right\},$$

where we used results of chapter III of Jacod and Shiryaev (1987), in particular sections 4 and 5.

We now calculate the Fisher information matrix corresponding to $X(T)$. It is the opposite of the expectation of the Hessian matrix of the log-likelihood.

$$I_T(\theta) = -E_\theta [H_\theta \log(f_T(X(T), \theta))].$$
The log-likelihood is
\[
\log(f_T(X(T), \theta)) = \log 2 + \frac{1}{2} \log(\det Q(\theta)) + \alpha T - \frac{1}{2}(X^T_0 Q(\theta) X_0 + X^T_T Q(\theta) X_T) - \frac{1}{2} \int_0^T \|Q(\theta) X_t\|^2 dt = \log 2 + \frac{1}{2} \log(\alpha^2 - \beta^2) + \alpha T - \frac{1}{2}(X^T_0 + X^T_T)(\alpha I + \beta A)(X_0 + X_T) - \frac{1}{2} \int_0^T \|Q(\theta) X_t\|^2 dt.
\]

And
\[
\|Q(\theta) X_t\|^2 = \alpha^2 \|X_t\|^2 + 2\alpha \beta X^T_0 A X_t + \beta^2 \|X_t\|^2.
\]

We deduce
\[
I_T(\theta) = -\frac{1}{2} E_\theta [H_\theta \log(\alpha^2 - \beta^2)] + \frac{T}{2} E_\theta [H_\theta \|Q(\theta) X_0\|^2] = I_0(\theta) + T I(\theta).
\]

With
\[
I(\theta) = \frac{1}{2} E_\theta \left[H_\theta \left(\alpha^2 \|X_0\|^2 + 2\alpha \beta X^T_0 A X_0 + \beta^2 \|X_0\|^2\right)\right].
\]

Using the expression of the variance of $X_T$ (2.4) we deduce,
\[
I(\theta) = Q(\theta)^{-1}.
\]

Let $T > 0$ and $h > 0$, we consider the problem of predicting $X_{T+h}$ given $X(T)$. Since $(X_t)_{t \geq 0}$ is a Markov process, it holds
\[
E_\theta[X_{T+h} | X(T)] = E_\theta[X_{T+h} | X_T] = r(X_T, \theta) = e^{-hQ(\theta)} X_T.
\]

The formula to differentiate a matrix exponential $e^{M(\theta)}$ with respect to $\theta$ is
\[
\frac{\partial e^{M(\theta)}}{\partial \theta_i} = \int_0^1 e^{uM(\theta)} \frac{\partial M(\theta)}{\partial \theta_i} e^{(1-u)M(\theta)} du.
\]

The partial derivatives of $r$ with respect to the components of $\theta$ are
\[
\frac{\partial}{\partial \alpha} r(X_T, \theta) = -he^{-hQ(\theta)} X_T,
\]
\[
\frac{\partial}{\partial \beta} r(X_T, \theta) = -he^{-hQ(\theta)} A X_T = -he^{-hQ(\theta)} A X_T.
\]
because $A$ and $e^{-hQ(\theta)}$ commute. Let

$$M_T = \begin{pmatrix} X_{1,T} & X_{2,T} \\ X_{2,T} & X_{1,T} \end{pmatrix} = X_{1,T}J_2 + X_{2,T}A$$

with $X_{1,T}$ and $X_{2,T}$ the two components of $X_T$. Then

$$J_\theta [r(X_T, \theta)] = -he^{-hQ(\theta)}M_T.$$

2.3.2. Asymptotic efficiency. Let

$$U_T = J_\theta [r(X_T, \theta)] = -he^{-hQ(\theta)}M_T.$$

To make notations lighter we note $U_T, M_T, X^{(1)}$ and $X^{(2)}$ independent copies of $U_T, M_T, X_{1,T}$ and $X_{2,T}$ respectively.

Let $\hat{\theta}_T$ be the MLE of $\theta$. One wants to predict $X_{T+h}$ with the plug-in predictor

$$\hat{r}_T(X_T) = r(X_T, \hat{\theta}_T) = e^{-hQ(\hat{\theta}_T)}X_T.$$ 

The MLE $\hat{\theta}_T$ satisfies

$$V_T = \sqrt{T}(\hat{\theta}_T - \theta) \xrightarrow{T \to \infty} V \sim N(0, I(\theta)^{-1}).$$

It holds

$$\nu_T(\theta) = \int_{\mathbb{R}^2} \left( J_\theta r(z, \theta) \right) I_T(\theta)^{-1} \left( J_\theta r(z, \theta) \right)' d\mu_\theta(z)$$

$$= E_\theta \left[ UI_T(\theta)^{-1}U' \right]$$

$$= E_\theta \left[ UI_0(\theta)^{-1}U' \right] + T \nu_*(\theta),$$

with

$$\nu_*(\theta) = E_\theta \left[ UI(\theta)^{-1}U' \right] = \int_{E} \left( J_\theta r(x, \theta) \right) I(\theta)^{-1} \left( J_\theta r(x, \theta) \right)' d\mu_\theta(x).$$

Then

$$\left\| \nu_T(\theta) - \nu_*(\theta) \right\|_{M_2} = o(\|\nu_T(\theta)\|_{M_2}).$$

We calculate $\nu_*(\theta)$

$$\nu_*(\theta) = h^2 e^{-hQ(\theta)}E_\theta \left[ MI(\theta)^{-1}M' \right] e^{-hQ(\theta)}.$$
and

$$E \rho \left[ (X(1) I_2 + X(2) A) (\alpha I_2 + \beta A) (X(1) I_2 + X(2) A) \right]$$

$$= E \rho \left[ (\alpha (X^2(1) + X^2(2)) + 2\beta X(1) X(2)) I_2 
+ (\beta (X^2(1) + X^2(2)) + 2\alpha X(1) X(2)) A \right]$$

$$= \frac{1}{2(\alpha^2 - \beta^2)} \begin{pmatrix}
2\alpha^2 - 2\beta^2 & 2\alpha\beta - 2\alpha^2 \\
2\alpha\beta - 2\alpha^2 & 2\alpha^2 - 2\beta^2
\end{pmatrix} = I_2,$$

hence $E \rho [M I(\theta)^{-1} M'] = I_2$, therefore

$$\nu_\theta(\theta) = h^2 e^{-2hQ(\theta)},$$

The QER with respect to $\mu_\theta = \mathcal{N}(0, Q(\theta)^{-1})$ is

$$\rho_T(\theta) = E \rho \left[ (e^{-hQ(\tilde{\theta}_T)} - e^{-hQ(\theta)}) E \rho (X_T X_T') (e^{-hQ(\tilde{\theta}_T)} - e^{-hQ(\theta)})' \right],$$

hence

$$\rho_T(\theta) = \frac{1}{2} E \rho \left[ (e^{-hQ(\tilde{\theta}_T)} - e^{-hQ(\theta)})Q(\theta)^{-1}(e^{-hQ(\tilde{\theta}_T)} - e^{-hQ(\theta)})' \right].$$

For all $\theta \in \Theta$, it holds

$$Q(\theta) = PD(\theta)P^{-1} \quad \text{and} \quad e^{-hQ(\theta)} = Pe^{-hD(\theta)}P^{-1}$$

with

$$P = \begin{pmatrix} 1 & 1 \\
1 & -1 \end{pmatrix} \quad \text{et} \quad D(\theta) = \begin{pmatrix} \alpha + \beta & 0 \\
0 & \alpha - \beta \end{pmatrix}.$$ 

The QER is

$$\rho_T(\theta) = \frac{1}{2} E \rho \left[ P(e^{-hD(\tilde{\theta}_T)} - e^{-hD(\theta)})P^{-1} Q(\theta)^{-1} P(e^{-hD(\tilde{\theta}_T)} - e^{-hD(\theta)})P^{-1} \right]$$

$$= \frac{P}{2} E \rho \left[ (e^{-hD(\tilde{\theta}_T)} - e^{-hD(\theta)})D(\theta)^{-1} (e^{-hD(\tilde{\theta}_T)} - e^{-hD(\theta)}) \right] \ P^{-1}$$

$$= \frac{P}{2} \begin{pmatrix}
\frac{\rho}{\alpha + \beta} & 0 \\
0 & \frac{\rho}{\alpha - \beta}
\end{pmatrix} e^{-hQ(\theta)} \ \text{P}^{-1}.$$ 

We consider the reparameterization

$$\xi = \begin{pmatrix} \eta \\
\gamma \end{pmatrix} = \begin{pmatrix} e^{-h(\alpha + \beta)} \\
e^{-h(\alpha - \beta)} \end{pmatrix},$$
then the MLE of $\xi$ is

$$\hat{\xi}_T = \begin{pmatrix} \tilde{\eta}_T \\ \tilde{\gamma}_T \end{pmatrix} = \begin{pmatrix} \exp(-h(\hat{\alpha}_T + \hat{\beta}_T)) \\ \exp(-h(\hat{\alpha}_T - \hat{\beta}_T)) \end{pmatrix}.$$ 

It holds

$$I(\xi)^{-1} = (J_\theta \xi(\theta))^{-1} (J_\theta \xi(\theta))^T,$$

$$J_\theta \xi(\theta) = -h \begin{pmatrix} \exp(-h(\alpha + \beta)) & \exp(-h(\alpha - \beta)) \\ \exp(-h(\alpha - \beta)) & -\exp(-h(\alpha + \beta)) \end{pmatrix},$$

hence

$$I(\xi)^{-1} = 2h^2 \begin{pmatrix} (\alpha + \beta) \exp(-2h(\alpha + \beta)) & 0 \\ 0 & (\alpha - \beta) \exp(-2h(\alpha - \beta)) \end{pmatrix}. $$

The estimator $\hat{\xi}_T$ satisfies

$$T \rho_T(\theta) \xrightarrow{T \to \infty} h^2 P \begin{pmatrix} \exp(-2h(\alpha + \beta)) & 0 \\ 0 & \exp(-2h(\alpha - \beta)) \end{pmatrix} \begin{pmatrix} e^{-hQ(\theta)}X_T \\ 0 \end{pmatrix} \Rightarrow P \begin{pmatrix} e^{-2hD(\theta)} \end{pmatrix},$$

hence

$$\rho_T(\theta) \xrightarrow{T \to \infty} h^2 e^{-2hQ(\theta)} = \nu_*(\theta).$$

Thus $r(\cdot, \hat{\theta}_T)$ is an asymptotically efficient estimator of $r(\cdot, \theta)$ for the QER with respect to $\mu_\theta = \mathcal{N}(0, \frac{Q(\theta)^{-1}}{2}).$

Applying Theorem 2.8 p.121 of Kutoyants (2004), for all $\theta \in \Theta$, there is a compact neighbourhood $V(\theta)$ of $\theta$, such that the following convergence holds uniformly over $V(\theta),$

$$\sqrt{T}(e^{-hQ(\theta)}X_T - e^{-hQ(\theta)}x) \Rightarrow \mathcal{N}(0, (J_\theta(e^{-hQ(\theta)}x))^{-1} (J_\theta(e^{-hQ(\theta)}x))^T),$$

where the convergence in distribution is taken with respect to $P_\theta^T$, the distribution of $X(T)$. Moreover the model is LAN (see Kutoyants (2004)). Hence applying Proposition 2.1, we deduce that the family of estimators $(e^{-hQ(\theta)}x, T \geq 0)$ is regular. Therefore Assumption 2.1 is fulfilled.

We conclude that the results pertaining to the estimation of the regression function we have seen in this section apply to the problem of forecasting of the bivariate stationary Ornstein-Uhlenbeck process that we consider here, and the plug-in estimator $r(\cdot, \hat{\theta}_T)$ is asymptotically efficient. We will see in Section 3 the extension of the results to the prediction problem, in particular we will verify Assumption 2.2.
3. Limits of risks for prediction and regression. In this section we give conditions under which the QEP and the QER are asymptotically equivalent. This asymptotic equivalence has been used in the previous section to deduce an asymptotic bound on QEP from an asymptotic bound on QER. Here we consider the plug-in estimators of the regression function, \( \hat{r}_T = r(\cdot, \hat{\theta}_T) \) where \( \hat{\theta}_T \) is an estimator of \( \theta \) based on \( X_T \), and we consider the corresponding predictors, of the form \( \hat{r}_T(X_T) = r(X_T, \hat{\theta}_T) \). The QER will always be taken with respect to \( \mu_\theta = P_{\theta, X_T} \) the distribution of \( X_T \).

3.1. Assumptions and lemma. The QEP of the predictor \( r(X_T, \hat{\theta}_T) \) is

\[
R_T(\theta) = E_\theta (r(X_T, \hat{\theta}_T) - r(X_T, \theta))^2, \quad \theta \in \Theta.
\]

In order to compare \( R_T(\theta) \) with \( \rho_T(\theta) \), we consider an auxiliary predictor \( r(X_T, \hat{\theta}_{S(T)}) \) and the corresponding estimator of the regression function \( r(\cdot, \hat{\theta}_{S(T)}) \) where \( \hat{\theta}_{S(T)} \) is based on \( X_{(S(T))} \) with \( S : \mathbb{R}^+ \rightarrow \mathbb{R}^+ \) a function such that \( S(T) \leq T \) for all \( T \geq 0 \) and \( S(T) \sim T \), as \( T \to \infty \). In what follows, \( S \) will always represent \( S(T) \), omitting the argument \( T \) in order to make notations lighter.

We introduce \( R_T^S(\theta) \) the QEP of the predictor \( r(X_T, \hat{\theta}_S) \), and \( \rho_T^S(\theta) \) the QER of the estimator \( r(\cdot, \hat{\theta}_S) \). These quantities are

\[
R_T^S(\theta) = E_\theta \left( r(X_T, \hat{\theta}_S) - r(X_T, \theta) \right)^2,
\]

\[
\rho_T^S(\theta) = \int_E E_\theta \left( r(x, \hat{\theta}_S) - r(x, \theta) \right)^2 d\mu_\theta(x).
\]

For all \( T \geq 0 \), we define the function \( \bar{\theta}_S : \mathcal{X}_S \rightarrow \Theta \), such that \( \bar{\theta}_S(X_{(S)}) = \hat{\theta}_S \), where \( \mathcal{X}_S \) is the space of the paths of \( X_{(S)} \) equipped with the smallest \( \sigma \)-field that makes the coordinate applications continuous. Let

\[
\Delta r(x, \xi) = r(x, \bar{\theta}_S(\xi)) - r(x, \theta), \quad \forall x \in E, \xi \in \mathcal{X}_S,
\]

where the dependence of \( \Delta r \) in \( T \) and \( \theta \) is left implicit, to make notations lighter.

The distribution \( P_{\theta,(X_T,X_{(S)})} \) of \( (X_T, X_{(S)}) \), is assumed to be dominated by a \( \sigma \)-finite measure \( \lambda \) and \( f_{\theta,X_T}, f_{\theta,X_{(S)}} \) and \( f_{\theta,(X_T,X_{(S)})} \) represent the densities of \( X_T, X_{(S)} \) and \( (X_T, X_{(S)}) \) respectively. We can then write \( R_T^S(\theta) \) and \( \rho_T^S(\theta) \) in the following way,

\[
R_T^S(\theta) = \int_{E \times \mathcal{X}_S} (\Delta r(x, \xi))^2 dP_{\theta,(X_T,X_{(S)})}(x, \xi)
\]

\[
= \int_{E \times \mathcal{X}_S} (\Delta r(x, \xi))^2 f_{\theta,(X_T,X_{(S)})} d\lambda(x, \xi),
\]
and

\[ \rho_T^S(\theta) = \int_{E \times X_S} (\Delta r(x, \xi))^2 dP_{\theta, X_T}(x) dP_{\theta, X(\xi)}(\xi) \]

\[ = \int_{E \times X_S} (\Delta r(x, \xi))^2 f_{\theta, X_T}(x) f_{\theta, X(\xi)}(\xi) d\lambda(x, \xi). \]

And let

\[ \Delta f(x, \xi) = |f_{\theta, (X_T, X(\xi))}(x, \xi) - f_{\theta, X_T}(x) f_{\theta, X(\xi)}(\xi)|, \]

letting the dependence of \( \Delta f \) in \( T \) and \( \theta \) implicit.

We measure the dependence of \( X_T \) and \( X(\xi) \) with the coefficient

\[ \bar{\beta}(S, T) = \int_{E \times X_S} (\Delta f) d\lambda. \]

**Remark 3.1.** This coefficient is bounded from above by the usual \( \beta \)-mixing coefficient. It holds

\[ \bar{\beta}(S, T) \leq 2\beta(T - S) \]

with

\[ \beta(t) = \sup_{s>0} \| P_{0, \theta}^s \otimes P_{s+t, \theta}^\infty - P_{s+t, \theta} \|_{TV} \]

where \( P_{0, \theta}^s \) is the distribution of \((X_u, 0 \leq u \leq s)\), \( P_{s+t, \theta}^\infty \) the distribution of \((X_u, u \geq s + t)\), \( P_{s, t, \theta} \) the joint distribution of \(( (X_u)_{0 \leq u \leq s}, (X_u)_{u \geq s+t}) \)

and \( \| \cdot \|_{TV} \) the total variation norm for signed measures, i.e. if \( \mu \) is a signed measure on a measurable space \( \mathcal{A} \), then \( \| \mu \|_{TV} = \sup_{A \in \mathcal{A}} | \mu(A) | \). Intuitively, the coefficient \( \bar{\beta}(S, T) \) is a way to quantity the dependence between \( X(\xi) \) and \( X_T \). A reference on mixing coefficients is Doukhan (1994). There are other approaches to measure the dependence between the past and the future, see Dedecker et al. (2007).

We now make the following assumption.

**Assumption 3.1.**

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

**Lemma 3.1.** Under Assumption 3.1, if the limit \( \lim_{T \to \infty} T \rho_T^S \) exists and is non zero then,

\[ \lim_{T \to \infty} T \rho_T^S = \lim_{T \to \infty} T \rho_T^S. \]

The proof is similar to the proof of Lemma 3.1 in Bosq and Onzon (2012).

We now state a lemma which will be useful for the proof of Theorem 3.1.
Lemma 3.2. Let $U$, $V$ and $W$ be column vectors in $\mathbb{R}^k$, then

$$\|(U - V)^\times - (W - V)^\times\|_{M_k} \leq \|U - W\|_{\mathbb{R}^k}^2 + 2\|U - W\|_{\mathbb{R}^k}\|W - V\|_{\mathbb{R}^k}$$

Proof.

\[
(U - V)^\times - (W - V)^\times = ((U - W) + (W - V))^\times - (W - V)^\times \\
= (U - W)^\times + (U - W)(W - V)' + (W - V)(U - W)'.
\]

Hence

$$\|(U - V)^\times - (W - V)^\times\|_{M_k} \leq \|(U - W)^\times\|_{M_k} + \|(U - W)(W - V)\|_{M_k} + \|(W - V)(U - W)\|_{M_k} \leq \|U - W\|_{\mathbb{R}^k}^2 + 2\|U - W\|_{\mathbb{R}^k}\|W - V\|_{\mathbb{R}^k}.$$

Assumption 3.2. For all $\theta \in \Theta$, $\lim_{T \to \infty} T\mathbb{E}_\theta \|\hat{\theta}_T - \hat{\theta}_S\|_{\mathbb{R}^d}^2 = 0$.

Remark 3.2. Proposition 3.1, which we will see further, gives conditions under which Assumption 3.2 is fulfilled.

Finally we make the following assumption.

Assumption 3.3. For all $\theta \in \Theta$, the following limits exist and are equal, and let $R(\theta) \in M_k$ such that,

$$\lim_{T \to \infty} T\rho_T^S(\theta) = \lim_{T \to \infty} T\rho_T(\theta) = R(\theta).$$

Remark 3.3. It is straightforward to see that Assumption 3.3 is fulfilled for our problem of forecasting of the bivariate stationary Ornstein-Uhlenbeck process for $S = T - \sqrt{T}$, replacing $\hat{\theta}_T$ by $\hat{\theta}_S$ in the derivation we made in Section 2.3.2 to calculate the limit of $\rho_T(\theta)$.

3.2. Asymptotic equivalence of QEP and QER.
3.2.1. **Theorem of asymptotic equivalence of the risks.** The two following theorems give conditions for asymptotic equivalence between QEP and QER. They are a generalization of Propositions 3.1 and 3.2 in Bosq and Onzon (2012) from the univariate to multivariate case.

**Theorem 3.1.** Under Assumptions 3.1, 3.2 and 3.3, if there exists a deterministic constant \( C > 0 \), telle que

\[
\|r(X_T, \theta^*) - r(X_T, \theta)\|_{\mathbb{R}^k} \leq C \|\theta^* - \theta\|_{\mathbb{R}^d}, \quad \forall \theta, \theta^* \in \Theta,
\]

then \( \lim_{T \to \infty} TR_T = \lim_{T \to \infty} T\rho_T \).

**Proof.** Assumption 3.3 entails

\[
\lim_{T \to \infty} T\rho_T^S(\theta) = \lim_{T \to \infty} T\rho_T(\theta).
\]

Assumption 3.1 and Lemma 3.1 give \( \lim_{T \to \infty} TR_T^S(\theta) = \lim_{T \to \infty} T\rho_T^S(\theta) \), hence

\[
\lim_{T \to \infty} TR_T^S(\theta) = \lim_{T \to \infty} T\rho_T(\theta).
\]

Condition (3.1) implies

\[
TE_\theta \|r(X_T, \hat{\theta}_T) - r(X_T, \hat{\theta}_S)\|_{\mathbb{R}^k}^2 \leq C^2 TE_\theta \|\hat{\theta}_T - \hat{\theta}_S\|_{\mathbb{R}^d}^2.
\]

Combining with Assumption 3.2 we obtain

\[
(3.2) \quad TE_\theta \|r(X_T, \hat{\theta}_T) - r(X_T, \hat{\theta}_S)\|_{\mathbb{R}^k}^2 \xrightarrow{T \to \infty} 0.
\]

Let \( U = r(X_T, \hat{\theta}_T) \), \( V = r(X_T, \theta) \) and \( W = r(X_T, \hat{\theta}_S) \), then using Lemma 3.2,

\[
\|T(R_T - R_T^S)\|_{M_k} = TE_\theta \|\varphi - \varphi^2\|_{M_k}^2
\]

\[
= TE_\theta \|\varphi - \varphi^2\|_{M_k}^2
\]

\[
\leq TE_\theta \|\varphi - \varphi^2\|_{M_k}^2
\]

\[
\leq TE_\theta \|\varphi - \varphi^2\|_{M_k}^2
\]

\[
= 2 \left( TE_\theta \|\varphi - \varphi^2\|_{M_k}^2 \right)^{1/2}.
\]

Using (3.2) we get \( TE_\theta \|U - W\|_{\mathbb{R}^k}^2 \xrightarrow{T \to \infty} 0 \). And moreover

\[
TE_\theta \|W - V\|_{\mathbb{R}^k}^2 = \text{trace}(TR_T^S(\theta)) \xrightarrow{T \to \infty} \text{trace}(R(\theta)).
\]

Therefore

\[
\|T(R_T(\theta) - R_T^S(\theta))\|_{M_k} \xrightarrow{T \to \infty} 0.
\]
Condition (3.1) is somewhat restrictive. For instance it is not fulfilled by our problem of forecasting of an Ornstein-Uhlenbeck process. The following result hinges upon less restrictive conditions.

**Theorem 3.2.** We assume the following conditions hold.

1. \( T^2 E_\theta \| \hat{\theta}_T - \hat{\theta}_S \|^4_{\mathbb{R}^d} = \mathcal{O}(1), \) as \( T \to \infty. \)
2. There is a measurable function \( \ell : E \to \mathbb{R}_+ \) such that for all \( \theta, \theta^* \in \Theta, \)
   \[ \| r(X_T, \theta^*) - r(X_T, \theta) \|_{\mathbb{R}^d} \leq \ell(X_T) \| \theta^* - \theta \|_{\mathbb{R}^d}. \]
3. \( \exists \nu > 0 \) such that 
   \[ C_\theta = \sup_T E_\theta (\ell^{4+\nu}(X_T)) < \infty. \]

Then, under Assumptions 3.1, 3.2 and 3.3, \( \lim_{T \to \infty} T R_T = \lim_{T \to \infty} T \rho_T. \)

We refer to the proof of Proposition 3.2 in Bosq and Onzon (2012), where the result is proved for \( \Theta \subset \mathbb{R} \) and \( r \) taking its values in \( \mathbb{R}, \) the proof makes use of Proposition 3.1 in Bosq and Onzon (2012), which is the restriction of Theorem 3.1 to the one-dimensional case. The generalization to the multidimensional case is straightforward, using Theorem 3.1 instead of Proposition 3.1 in Bosq and Onzon (2012).

3.2.2. Verification of Assumption 3.2. We now give a result that allows to verify Assumption 3.2. We make the following assumption.

**Assumption 3.4.** Let \( \theta \in \Theta. \) Assume the family \( (P^T_\theta, \theta \in \Theta), T \geq 0, \) satisfies the following conditions.

1. The family \( (P^T_\theta, \theta \in \Theta), T \geq 0, \) is LAN.
2. For all \( \eta > 0, \) \( \lim_{T \to \infty} P^T_\theta \left( \| \Delta_T(\theta) - \frac{\sqrt{T}}{\sqrt{S}} \Delta_S(\theta) \|^2_{\mathbb{R}^d} > \eta \right) = 0. \)

The family of estimators \( (\hat{\theta}_T, T \geq 0), \) satisfies the following conditions.

3. The family \( (\hat{\theta}_T, T \geq 0), \) is regular at \( \theta. \)
4. \( P^T_\theta \circ \sqrt{T}(\hat{\theta}_T - \theta) \xrightarrow{T \to \infty} \mathcal{N}(0, I(\theta)^{-1}) \)
5. \( \lim_{T \to \infty} T E_\theta \| \hat{\theta}_T - \theta \|^2_{\mathbb{R}^d} = \text{trace}(I(\theta)^{-1}) \)

**Proposition 3.1.** Let \( \theta \in \Theta. \) Under Assumption 3.4,

\[ \lim_{T \to \infty} T E_\theta \| \hat{\theta}_T - \bar{\theta}_S \|^2_{\mathbb{R}^d} = 0. \]

**Proof.** Let \( \bar{\theta}_T = \frac{1}{2}(\hat{\theta}_T + \bar{\theta}_S). \) For all vectors \( x \) and \( y, \) the parallelogram identity is

\[ \| x + y \|^2 + \| x - y \|^2 = 2\| x \|^2 + 2\| y \|^2. \]
Taking $x = \hat{\theta}_T - \theta$ and $y = \hat{\theta}_S - \theta$ in the normed vector space $L^2(P_\theta)$ we get

$$E_\theta\|\hat{\theta}_T + \hat{\theta}_S - 2\theta\|^2_{\mathbb{R}^d} + E_\theta\|\hat{\theta}_T - \hat{\theta}_S\|^2_{\mathbb{R}^d} = 2E_\theta\|\hat{\theta}_T - \theta\|^2_{\mathbb{R}^d} + 2E_\theta\|\hat{\theta}_S - \theta\|^2_{\mathbb{R}^d}.$$ 

Hence

$$4T E_\theta\|\hat{\theta}_T - \theta\|^2_{\mathbb{R}^d} + TE_\theta\|\hat{\theta}_S - \theta\|^2_{\mathbb{R}^d} = 2TE_\theta\|\hat{\theta}_T - \theta\|^2_{\mathbb{R}^d} + 2TE_\theta\|\hat{\theta}_S - \theta\|^2_{\mathbb{R}^d}.$$ 

To complete the proof it remains to prove that

$$\lim_{T \to \infty} TE_\theta\|\hat{\theta}_T - \theta\|^2_{\mathbb{R}^d} \geq \text{trace}(I(\theta)^{-1}).$$

From Proposition 2.2, the normalized error of estimators $\hat{\theta}_T$ and $\hat{\theta}_S$ have the following asymptotic expansions.

$$\sqrt{T}(\hat{\theta}_T - \theta) = I(\theta)^{-1}\Delta_T + o_{p,T}(1),$$

$$\sqrt{S}(\hat{\theta}_S - \theta) = I(\theta)^{-1}\Delta_S + o_{p,T}(1).$$

Where we omit the argument $\theta$ of $\Delta_T$ and $\Delta_S$ to make notations lighter. We deduce,

$$\sqrt{T}(\hat{\theta}_T - \theta) = \frac{1}{2}\sqrt{T}(\hat{\theta}_T - \theta) + \frac{1}{2}\sqrt{S}(\hat{\theta}_S - \theta)$$

$$= \frac{1}{2}I(\theta)^{-1}\left(\Delta_T + \sqrt{\frac{T}{S}}\Delta_S\right) + o_{p,T}(1)$$

$$= I(\theta)^{-1}\left(\Delta_T + \frac{1}{2}\left(\sqrt{T}\Delta_S - \Delta_T\right)\right) + o_{p,T}(1).$$

Yet $P_\theta^T \circ \Delta_T \Rightarrow \mathcal{N}(0, I(\theta))$ and $\frac{\sqrt{T}}{\sqrt{S}}\Delta_S - \Delta_T$ converges toward 0 in probability. Hence, applying Slutsky’s lemma we get,

$$P_\theta^T \circ \sqrt{T}(\hat{\theta}_T - \theta) \overset{T \to \infty}{\Rightarrow} \mathcal{N}(0, I(\theta)^{-1}).$$

Therefore, from Lemma 2.2,

$$\lim_{T \to \infty} TE_\theta\|\hat{\theta}_T - \theta\|^2_{\mathbb{R}^d} \geq \text{trace}(I(\theta)^{-1}).$$

This completes the proof. \qed
3.2.3. Bivariate Ornstein-Uhlenbeck process. We have seen in Remark 3.3 that Assumption 3.3 is fulfilled for our problem of forecasting of the bivariate stationary Ornstein-Uhlenbeck process. We are now going to see that Assumptions 3.1 and 3.2, and conditions of Theorem 3.2 are fulfilled too.

We begin with Assumption 3.1. We take $S = T - \sqrt{T}$, and $\lambda = \ell \otimes \nu$ with $\ell$ Lebesgue’s measure over $\mathbb{R}$ and $\nu$ the distribution of the process $(U_t + W_t, t \geq 0)$, where $(W_t, t \geq 0)$ is a standard bivariate Wiener process and $U \sim \mathcal{N}(0, I_2)$ independent (with $I_2$ the unit matrix of $\text{M}_2$). It holds

$$\|\Delta_r(x, \xi)\|_\mathbb{R}^2 = \|\left(e^{-Q(\theta T)^h} - e^{-Q(\theta)^h}\right)x\|_\mathbb{R}^2 \leq 2\|x\|_\mathbb{R}^2, \quad \forall x \in \mathbb{R}^2.$$ 

We deduce that the first condition of Assumption 3.1 is fulfilled,

$$\|\Delta r\|_{L^m(\Delta f \lambda)} \leq \int \|\Delta r\|_m^m \Delta f d\lambda \leq 4E_\theta\|X_0\|^m \|x\|_m < \infty,$$

where $\| \cdot \|_m$ is the $m$ norm of $\mathbb{R}^2$. In particular for $m = 4$, this bound is

$$4E_\theta\|X_0\|_4^4 = 12((E_\theta X_0^2, 1^2) + (E_\theta X_0^2, 2^2)) = \frac{12\alpha^2}{(\alpha^2 - \beta^2)^2}.$$

The multivariate Ornstein-Uhlenbeck process is geometrically $\beta$-mixing (see Veretennikov (1987)), which allows to conclude that the second condition of Assumption 3.1 is fulfilled for all $m > 2$.

We are now going to see that Assumption 3.4 is verified, which will imply Assumption 3.2. We have already seen that conditions 1, 3, 4 et 5 are fulfilled, we turn to condition 2.

The family $(P^T_\theta, \theta \in \Theta)$, $T > 0$, is LAN (see Kutoyants (2004) p.113) with,

$$\Delta_T(\theta) = T^{-1/2} \int_0^T (J_\theta S(\theta, X_t))' dW_t,$$

where $S(\theta, X_t) = -Q(\theta)X_t$. Therefore $\Delta_T(\theta) = T^{-1/2} \int_0^T Q(X_t)dW_t$. Now

$$\Delta_T - \sqrt{\frac{T}{S}} \Delta S = -T^{-1/2} \int_0^T Q(X_t)dW_t + T^{1/2} S^{-1} \int_0^S Q(X_t)dW_t$$

$$= T^{-1/2} \left(\frac{T}{S} - 1\right) \int_0^S Q(X_t)dW_t + T^{-1/2} \int_S^T Q(X_t)dW_t$$

$$= A_t + B_t.$$ 

$$E_\theta\|A_t\|^2 \|_\mathbb{R}^2 = T^{-1} \left(\frac{T}{S} - 1\right)^2 \int_0^S E_\theta Q(X_t)^2 dt$$

$$= \frac{S}{T} \left(\frac{T}{S} - 1\right)^2 Q(\theta)^{-1} \xrightarrow{T \to \infty} 0.$$
The convergence of $A_t$ to 0 in $L^2$ implies its convergence in probability.

\[
E_{\theta} \| B_t \|^2_{L^2} = T^{-1} \int_S^T E_{\theta} Q(X_t)^2 dt = \frac{T - S}{T} Q(\theta)^{-1} = \left(1 - \frac{S}{T}\right) Q(\theta)^{-1} \xrightarrow{T \to \infty} 0.
\]

Hence $B_t$ converges to 0 in probability. We deduce

\[
\left\| \Delta T - \frac{\sqrt{T}}{\sqrt{S}} \Delta S \right\|_{L^2} \leq \| A_t \|_{L^2} + \| B_t \|_{L^2} \xrightarrow{T \to \infty} 0.
\]

Therefore condition 2 of Assumption 3.4 is fulfilled.

Condition 1 of Theorem 3.2 is fulfilled from Theorem 2.8 p.121 of Kutoyants (2004). We are now going to see that condition 2 is fulfilled too.

\[
\| r(x, \theta) - r(x, \theta^*) \|_{L^2} = \| \partial_{\theta} r(x, \tilde{\theta})(\theta - \theta^*) \|_{L^2} \\
\leq \| \partial_{\theta} r(x, \tilde{\theta}) \|_{M_2} \| \theta - \theta^* \|_{L^2} \\
\leq \frac{h}{2} \| P \|_{M_2} \| e^{-hD(\tilde{\theta})} \|_{M_2} \| M(x) \|_{M_2} \| \theta - \theta^* \|_{L^2} \\
\leq \sqrt{2h} \| P \|_{M_2} \| e^{-hD(\tilde{\theta})} \|_{M_2} \| x \|_{L^2} \| \theta - \theta^* \|_{L^2} \\
\leq \ell(x) \| \theta - \theta^* \|_{L^2},
\]

with $\tilde{\theta} = \lambda \theta + (1 - \lambda) \theta^*$, $\lambda \in [0, 1]$, and $\ell(x) = \sqrt{2h} \| P \|_{M_2} \| e^{-hD(\tilde{\theta})} \|_{M_2} \| x \|_{L^2}$. For all $T \geq 0$, $X_T \sim N\left(0, Q(\theta)^{-1}\right)$ hence $\ell$ verifies condition 3 of Theorem 3.2, for all $\nu > 0$.

In conclusion, Assumption 2.2 is verified. The results pertaining to prediction of the previous section applies to the problem of forecasting of the bivariate stationary Ornstein-Uhlenbeck process that we consider here. The plug-in predictor $r(X_T, \tilde{\theta}_T)$ is asymptotically efficient.

3.3. Break down of the limiting QER. In this section we see results which give an expression of the limiting QER under some conditions. As a by-product we obtain the fulfillment of Assumption 3.3. We shall use the following notation, for all function $f \in L^2(\mu_\theta)$,

\[
\| f \|^2_{\mu_\theta} = \int_E \| f(x) \|^2 d\mu_\theta(x).
\]

In the univariate case, the limiting QER may take the following product form,

\[
\lim_{T \to \infty} T \rho_T(\theta) = \| \partial_{\theta} r(\cdot, \theta) \|^2_{\mu_\theta} \nu(\theta),
\]
with \( V(\theta) = \lim_{T \to \infty} T E_{\theta}(\hat{\theta}_T - \theta)^2 \). The result generalizes to the multivariate case with the following forms,

\[
\lim_{T \to \infty} T \rho_T(\theta) = \lim_{T \to \infty} T \int_E (J_{\theta r}(x, \theta)) E_{\theta}(\hat{\theta}_T - \theta)^2 (J_{\theta r}(x, \theta))' d\mu_\theta(x),
\]

\[
\lim_{T \to \infty} T \rho_T(\theta) = \int_E (J_{\theta r}(x, \theta)) V(\theta) (J_{\theta r}(x, \theta))' d\mu_\theta(x),
\]

with \( V(\theta) = \lim_{T \to \infty} T E_{\theta}(\hat{\theta}_T - \theta)^2 \). As a corollary we will deduce that Assumption 3.3 is verified, namely

\[
\lim_{T \to \infty} T \rho_T(\theta) = \lim_{T \to \infty} T \rho_T(\hat{\theta}_T). 
\]

### 3.3.1. Univariate case.

**Assumption 3.5.**

1. \( \exists \alpha \in (0,1], \exists c(x) \) such that \( \|c\|_{\mu_\theta} < \infty \) and

\[
|\partial_{\theta} r(x, \theta') - \partial_{\theta} r(x, \theta)| \leq c(x)|\theta' - \theta|^\alpha, \quad \forall x \in E, \forall \theta, \theta' \in \Theta, \forall T > 0.
\]

Moreover, \( \hat{\theta}_T \) is an estimator of \( \theta \) such that

3. \( \exists V(\theta) \) such that \( \lim_{T \to \infty} T E_{\theta}(\hat{\theta}_T - \theta)^2 = V(\theta) \),

4. \( T E_{\theta}|\theta_T - \theta|^{2+2\alpha} = o(1) \).

**Proposition 3.2.** Under Assumption 3.5, for all \( \theta \in \Theta \),

\[
\lim_{T \to \infty} T \rho_T(\theta) = \|\partial_{\theta} r(\cdot, \theta)\|_{\mu_\theta}^2 V(\theta).
\]

**Proof.**

\[
r(x, \hat{\theta}_T) - r(x, \theta) = (\hat{\theta}_T - \theta) \partial_{\theta} r(x, \hat{\theta}_T)
\]

where \( \hat{\theta}_T \in [\min(\theta, \hat{\theta}_T), \max(\theta, \hat{\theta}_T)] \). Soit

\[
\delta_T(x) = (\hat{\theta}_T - \theta)(\partial_{\theta} r(x, \hat{\theta}_T) - \partial_{\theta} r(x, \theta)).
\]

Hence

\[
r(x, \hat{\theta}_T) - r(x, \theta) = (\hat{\theta}_T - \theta) \partial_{\theta} r(x, \theta) + \delta_T(x),
\]

and using condition 1 of Assumption 3.5

\[
|\delta_T(x)| \leq c(x)|\hat{\theta}_T - \theta|^{1+\alpha}. \tag{3.3}
\]
Besides,
\[
\rho_T(\theta) = E_\theta(\hat{\theta}_T - \theta)^2 \int_E (\partial_\theta r(x, \theta))^2 d\mu_\theta(x) + \int_E E_\theta(\delta_T^2(x)) d\mu_\theta(x) \\
+ 2 \int_E E_\theta[(\hat{\theta}_T - \theta)\delta_T(x)] \partial_\theta r(x, \theta) d\mu_\theta(x) \\
= J_1 + J_2 + J_3.
\]
From condition (3.3) it ensues
\[
|J_2| \leq E_\theta(|\hat{\theta}_T - \theta|^{2+2\alpha}) \int_E c^2 d\mu_\theta.
\]
Since ||c||_{\mu_\theta} is bounded, condition 4 of Assumption 3.5 implies |J_2| = o\left(\frac{1}{T}\right).
Now (3.3) gives
\[
|J_3| \leq 2E_\theta(|\hat{\theta}_T - \theta|^{2+\alpha}) \int_E c(x)|\partial_\theta r(x, \theta)| d\mu_\theta(x) \\
\leq 2E_\theta|\hat{\theta}_T - \theta|^{2+\alpha}||c||_{\mu_\theta}||\partial_\theta r(\cdot, \theta)||_{\mu_\theta} \\
\leq 2E_\theta\left[|\hat{\theta}_T - \theta|^{2+\alpha}||c||_{\mu_\theta}||\partial_\theta r(\cdot, \theta)||_{\mu_\theta}\right] \\
\leq 2\left(\frac{1}{\sqrt{T}}\right) o\left(\frac{1}{\sqrt{T}}\right) = o\left(\frac{1}{T}\right).
\]
Finally \(TJ_1 \rightarrow \|\partial_\theta r(\cdot, \theta)\|_{\mu_\theta}^2 V(\theta)\).

We deduce that under conditions of Proposition 3.2, Assumption 3.3 is verified.

**Corollary 3.1.** Under Assumption 3.5, for all \(\theta \in \Theta\),
\[
\lim_{T \rightarrow \infty} T \rho_T(\theta) = \lim_{T \rightarrow \infty} T \rho_T^S(\theta) = ||\partial_\theta r(\cdot, \theta)||_{\mu_\theta}^2 V(\theta).
\]

**Proof.** If \(\hat{\theta}_T\) satisfies Assumption 3.5 then \(\hat{\theta}_S\) also satisfies this assumption because \(S \sim T\). Hence \(T \rightarrow \infty \lim T \rho_T^S(\theta) = ||\partial_\theta r(\cdot, \theta)||_{\mu_\theta}^2 V(\theta)\). \(\Box\)

**Remark 3.4.** When Proposition 3.2 applies, the asymptotic efficiency of a plug-in estimator \(r(\cdot, \hat{\theta}_T)\), or a plug-in predictor \(r(X_T, \hat{\theta}_T)\), comes down to the asymptotic efficiency of the estimator \(\hat{\theta}_T\) unless Assumption 3.5 holds and if \(V(\theta) \neq 0\), then \(r(\cdot, \hat{\theta}_T)\) is asymptotically efficient iff
\[
\lim_{T \rightarrow \infty} TE_\theta(\hat{\theta}_T - \theta)^2 = I(\theta)^{-1},
\]
assuming the family is LAN and $I(\theta)$ is the asymptotic Fisher information.

If in addition Assumptions 3.1 and 3.2 are satisfied, then

$$\lim_{T \to \infty} TR_T(\theta) = \|\partial_\theta r(\cdot, \theta)\|^2_{\mu_\theta} V(\theta),$$

and a plug-in predictor $r(X_T, \hat{\theta}_T)$ is asymptotically efficient iff

$$\lim_{T \to \infty} T E_\theta(\hat{\theta}_T - \theta)^2 = I(\theta)^{-1}.$$

3.3.2. **Multivariate case.** We now see generalizations of these results to the multivariate case.

**Assumption 3.6.**

1. $\exists \alpha \in (0, 1], \exists c: \mathbb{R}^k \to \mathbb{R}$ with $\|c\|_{\mu_\theta} < \infty$ such that

$$\|J_\theta r(x, \theta') - J_\theta r(x, \theta)\|_{M_{k,d}} \leq c(x)\|\theta'-\theta\|_{\mathbb{R}^d}^\alpha, \quad x \in E, \theta, \theta' \in \Theta.$$

Moreover, $\hat{\theta}_T$ is an estimator of $\theta$ such that

2. $T E_\theta\|\hat{\theta}_T - \theta\|^2_{\mathbb{R}^d} = O(1)$.
3. $T E_\theta\|\hat{\theta}_T - \theta\|^{3+2\alpha}_{\mathbb{R}^d} = o(1)$.
4. The following limit exists,

$$R(\theta) = \lim_{T \to \infty} T \int_E (J_\theta r(x, \theta)) E_\theta(\hat{\theta}_T - \theta)^2 (J_\theta r(x, \theta))^\prime d\mu_\theta(x).$$

**Proposition 3.3.** Under Assumption 3.6, for all $\theta \in \Theta$,

$$\lim_{T \to \infty} TR_T(\theta) = R(\theta).$$

**Proof.**

$$r(x, \hat{\theta}_T) - r(x, \theta) = (J_\theta r(x, \hat{\theta}_T))(\hat{\theta}_T - \theta),$$

where $\tilde{\theta}_T = \lambda \theta + (1 - \lambda)\hat{\theta}_T$ for some $\lambda \in [0, 1]$. Let

$$\delta_T(x, \theta) = (J_\theta r(x, \tilde{\theta}_T) - J_\theta r(x, \theta))(\hat{\theta}_T - \theta).$$

Then

$$r(x, \hat{\theta}_T) - r(x, \theta) = (J_\theta r(x, \theta))(\hat{\theta}_T - \theta) + \delta_T(x, \theta).$$

Thus

$$\rho_T(\theta) = J_1 + J_2 + J_3,$$
with

\[
J_1 = \int \left( J_\theta r(x, \theta) \right) E_{\theta}(\hat{\theta}_T - \theta)^2 \left( J_\theta r(x, \theta) \right)' d\mu_\theta(x),
\]

\[
J_2 = \int \left( \delta_T(x, \theta) \delta_T'(x, \theta) \right) d\mu_\theta(x),
\]

\[
J_3 = \int \left( \left( (J_\theta r(x, \theta)) (\hat{\theta}_T - \theta) \delta_T'(x, \theta) \right) + \delta_T(x, \theta)(\hat{\theta}_T - \theta)' (J_\theta r(x, \theta))' \right) d\mu_\theta(x).
\]

Condition 1 of Assumption 3.6 implies

\[
\|\delta_T(x, \theta)\|_{\mathbb{R}^k} \leq c(x) \|\hat{\theta}_T - \theta\|_{\mathbb{R}^d}^{1+\alpha}.
\]

Hence

\[
\|J_2\|_{M_k} \leq E_{\theta}(\|\hat{\theta}_T - \theta\|_{\mathbb{R}^d}^{2+2\alpha}) \|c\|_{\mu_\theta}^2.
\]

Then condition 3 of Assumption 3.6 implies

\[
\|J_2\|_{M_k} = o\left(\frac{1}{T}\right).
\]

Now,

\[
\|J_3\|_{M_k} \leq 2 \int \|J_\theta r(x, \theta)\|_{M_k,d} E_{\theta}[\|\hat{\theta}_T - \theta\|_{\mathbb{R}^d} \|\delta_T(x, \theta)\|_{\mathbb{R}^k}] d\mu_\theta(x).
\]

Using (3.4) it ensues,

\[
\|J_3\|_{M_k} \leq 2 E_{\theta}[\|\hat{\theta}_T - \theta\|_{\mathbb{R}^d}^{2+2\alpha} \|c\|_{\mu_\theta} \|\hat{\theta}_T - \theta\|_{\mathbb{R}^d}^{1+\alpha} \|\delta_T(x, \theta)\|_{\mathbb{R}^k}] d\mu_\theta(x)
\]

\[
\leq 2 E_{\theta}[\|\hat{\theta}_T - \theta\|_{\mathbb{R}^d}^{2+2\alpha} \|\hat{\theta}_T - \theta\|_{\mathbb{R}^d}^{1+\alpha} \|\delta_T(x, \theta)\|_{\mathbb{R}^k}] d\mu_\theta(x)
\]

\[
\leq 2 \left( E_{\theta}[\|\hat{\theta}_T - \theta\|_{\mathbb{R}^d}^2 \|\hat{\theta}_T - \theta\|_{\mathbb{R}^d}^{2+2\alpha} \|\delta_T(x, \theta)\|_{\mathbb{R}^k}^\frac{1}{2} \|\delta_T(x, \theta)\|_{\mathbb{R}^k}^\frac{1}{2}] \|\delta_T(x, \theta)\|_{\mathbb{R}^k} \|\hat{\theta}_T - \theta\|_{\mathbb{R}^d} \|\delta_T(x, \theta)\|_{\mathbb{R}^k} \|\delta_T(x, \theta)\|_{\mathbb{R}^k} \right)
\]

\[
\leq o\left(\frac{1}{T^{\frac{1}{2}}}\right) o\left(\frac{1}{T^{\frac{1}{2}}}\right) = o\left(\frac{1}{T}\right).
\]

Finally, from condition 4 of Assumption 3.6, \(TJ_1 \xrightarrow{T \to \infty} R(\theta)\).

We deduce that under the conditions of Proposition 3.3, Assumption 3.3 is satisfied.
Corollary 3.2. Under Assumption 3.6, for all $\theta \in \Theta$,

$$\lim_{T \to \infty} T \rho_T(\theta) = \lim_{T \to \infty} T \rho_T^S(\theta) = R(\theta).$$

Proof. The proof is similar to the univariate case. If $\hat{\theta}_T$ satisfies Assumption 3.6 then $\hat{\theta}_S$ also satisfies this assumption because $S \sim T$. Hence

$$\lim_{T \to \infty} T \rho_T^S(\theta) = R(\theta).$$

The asymptotic efficiency of a plug-in estimator or a plug-in predictor may come down to the asymptotic efficiency of the estimator of the parameter, as in the univariate case. For this we have to make the following additional assumption.

Assumption 3.7.

1. $\exists \alpha \in (0, 1], \exists C : \mathbb{R}^k \to \mathbb{R}$ with $\|c\|_{\mu_{\theta}} < \infty$ such that
   $$\|J_{\theta}(x, \theta') - J_{\theta}(x, \theta)\|_{M_{k,d}} \leq C(x)\|\theta' - \theta\|_{R^d}, \quad x \in E, \theta, \theta' \in \Theta.$$

2. For all $\theta \in \Theta$, there is a measurable function $g_{\theta} : E \to \mathbb{R}^+$ such that $\int_E g_{\theta}(x)d\mu_{\theta}(x) < \infty$ and, for $\mu_{\theta}$-almost all $x$, for all $T \geq 0$,
   $$\|J_{\theta}(x, \theta)\|^2_{M_{k,d}} \leq g_{\theta}(x).$$

Moreover, $\hat{\theta}_T$ is an estimator of $\theta$ such that, for all $\theta \in \Theta$,

3. $\exists V(\theta) \in M_d, \text{ such that } \lim_{T \to \infty} T E_{\theta}(\hat{\theta}_T - \theta)^2 = V(\theta)$.

4. $T E_{\theta}\|\hat{\theta}_T - \theta\|^2_{R^d} = o(1)$.

When Assumption 3.7 is satisfied, we let

$$R(\theta) = \int_E J_{\theta}(x, \theta)V(\theta)(J_{\theta}(x, \theta))^t d\mu_{\theta}(x).$$

Proposition 3.4. Under Assumption 3.7, for all $\theta \in \Theta$,

$$\lim_{T \to \infty} T \rho_T(\theta) = \int_E J_{\theta}(x, \theta)V(\theta)(J_{\theta}(x, \theta))^t d\mu_{\theta}(x).$$
**Proof.** We are going to apply Lebesgue’s dominated convergence theorem, as $T \to \infty$, to the integral

$$A_T = T \int_E (J_\theta r(x, \theta))E_\theta(\hat{\theta}_T - \theta) \times 2 \left( J_\theta r(x, \theta) \right)' d\mu_\theta(x).$$

Let $y \in \mathbb{R}^k$, for $\mu_\theta$-almost all $x$,

$$y'\left( J_\theta r(x, \theta) \right)E_\theta(\hat{\theta}_T - \theta) \times 2 \left( J_\theta r(x, \theta) \right)' y \xrightarrow{T \to \infty} J_\theta r(x, \theta)V(\theta)(J_\theta r(x, \theta))'.$$

From condition 3 of Assumption 3.7, $TE_\theta(\hat{\theta}_T - \theta) \times 2$ converges as $T \to \infty$ hence it is bounded. Let

$$C = \sup_{T \geq 0} \|TE_\theta(\hat{\theta}_T - \theta) \times 2\|_{M_d},$$

then, for $\mu_\theta$-almost all $x$,

$$|y'\left( J_\theta r(x, \theta) \right)E_\theta(\hat{\theta}_T - \theta) \times 2 \left( J_\theta r(x, \theta) \right)' y| \leq \|y\|^2_{\mathbb{R}^k} Cg_\theta(x).$$

Yet $\int_E g_\theta(x)d\mu_\theta(x) < \infty$, hence Lebesgue’s dominated convergence theorem implies

$$\lim_{T \to \infty} y' A_T y = y' \int_E J_\theta r(x, \theta)V(\theta)(J_\theta r(x, \theta))' d\mu_\theta(x)y.$$

This is true for all $y \in \mathbb{R}^k$ hence

$$\lim_{T \to \infty} A_T = \int_E J_\theta r(x, \theta)V(\theta)(J_\theta r(x, \theta))' d\mu_\theta(x).$$

So condition 4 of Assumption 3.6 is fulfilled with

$$R(\theta) = \int_E J_\theta r(x, \theta)V(\theta)(J_\theta r(x, \theta))' d\mu_\theta(x).$$

Now applying Proposition 3.3, we deduce the result. 

**Remark 3.5.** When Proposition 3.4 applies, the asymptotic efficiency of a plug-in estimator $r(\cdot, \hat{\theta}_T)$, or a plug-in predictor $r(X_T, \hat{\theta}_T)$, comes down to the asymptotic efficiency of the estimator $\hat{\theta}_T$. More precisely, if Assumption 3.5 is satisfied, the family is LAN and $I(\theta)$ is the asymptotic Fisher information, then $r(\cdot, \hat{\theta}_T)$ is asymptotically efficient iff

$$\lim_{T \to \infty} TE_\theta(\hat{\theta}_T - \theta) \times 2 = I(\theta)^{-1}.$$
If in addition Assumptions 3.1 and 3.2 are satisfied, then
\[ \lim_{T \to \infty} TR_T(\theta) = R(\theta), \]
and a plug-in predictor \( r(X_T, \hat{\theta}_T) \) is asymptotically efficient iff
\[ \lim_{T \to \infty} T E_\theta (\hat{\theta}_T - \theta)^2 = I(\theta)^{-1}. \]

3.3.3. Bivariate Ornstein-Uhlenbeck process. We are going to see that Assumptions 3.6 and 3.7 are satisfied for the problem of forecasting of the bivariate stationary Ornstein-Uhlenbeck process seen earlier.

Let \( \theta, \theta' \in \Theta \), and
\[ M(x) = \begin{pmatrix} x_1 & x_2 \\ x_2 & x_1 \end{pmatrix}, \]
then
\[
\| J_\theta r(x, \theta') - J_\theta r(x, \theta) \|_{M_2}^2 = h^2 \| (e^{-hQ(\theta')} - e^{-hQ(\theta)}) M(x) \|_{M_2}^2 \\
= h^2 \left( \| (e^{-hQ(\theta')} - e^{-hQ(\theta)}) x \|_{R_2}^2 + \| (e^{-hQ(\theta')} - e^{-hQ(\theta)}) Ax \|_{R_2}^2 \right) \\
= h^2 \left( \| r(x, \theta') - r(x, \theta) \|_{R_2}^2 + \| r(Ax, \theta') - r(Ax, \theta) \|_{R_2}^2 \right) \\
\leq h^2 \left( \| J_\theta r(x, \theta_1) \|_{M_2}^2 + \| J_\theta r(Ax, \theta_2) \|_{M_2}^2 \right) \| \theta' - \theta \|_{R_2}^2
\]
with \( \theta_1 = \lambda_1 \theta + (1 - \lambda_1) \theta^* \) and \( \theta_2 = \lambda_2 \theta + (1 - \lambda_2) \theta^* \) and \( \lambda_1, \lambda_2 \in [0, 1] \). Let
\[ c(x) = \sqrt{\| J_\theta r(x, \theta_1) \|_{M_2}^2 + \| J_\theta r(Ax, \theta_2) \|_{M_2}^2}. \]

Then
\[
\| c \|_{\mu_\theta}^2 = \| J_\theta r(\cdot, \theta_1) \|_{\mu_\theta}^2 + \| J_\theta r(\cdot, \theta_2) \|_{\mu_\theta}^2 \\
= h^2 \left( \text{trace} \left( Q(\theta)^{-1} e^{-2hQ(\theta_1)} \right) + \text{trace} \left( Q(\theta)^{-1} e^{-2hQ(\theta_2)} \right) \right) < \infty
\]
Hence condition 1 of Assumption 3.6 is fulfilled with \( \alpha = 1 \). The other conditions of Assumption 3.6 are verified applying Theorem 2.8 p.121 Kutoyants (2004). Condition 2 of Assumption 3.7 is fulfilled because \( \mu_\theta \) and \( r \) do not depend on \( T \) and \( E_\theta \| J_\theta r(X_T, \theta) \|_{M_2} < \infty \). Condition 3 is fulfilled because
\[ \lim_{T \to \infty} T E_{\theta} (\hat{\theta}_T - \theta)^2 = I(\theta)^{-1} \] (Theorem 2.8 p.121 of Kutoyants (2004)).

The other conditions of Assumptions 3.7 are shared with Assumption 3.6. So we can apply Theorems 3.3 and 3.4 to this problem, and we find again

\[ \lim_{T \to \infty} T \rho_T (\theta) = E_{\theta} (U I(\theta)^{-1} U') , \]

with \( U = J_{\theta} r(X_0, \theta) \).

**Acknowledgements.** I wish to thank Professors Denis Bosq and Delphine Blanke for their helpful comments during the preparation of this paper. I thank an anonymous referee for helpful comments that made possible to improve the article.

**References.**

Akritas, M. G. and Johnson, R. A. (1982). Efficiencies of tests and estimators for \( p \)-order autoregressive processes when the error distribution is nonnormal. *Ann. Inst. Statist. Math.* 34 579–589. MR695076 (84g:62139)

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1998). *Efficient and adaptive estimation for semiparametric models*. Springer-Verlag, New York. Reprint of the 1993 original. MR1623559 (99c:62076)

Bosq, D. (2007). Sufficiency and efficiency in statistical prediction. *Statist. Probab. Lett.* 77 280–287. MR2339031

Bosq, D. (2010). Exact asymptotic bias for estimators of the Ornstein-Uhlenbeck process. *Stat. Inference Stoch. Process.* 13 133–145. MR2653984

Bosq, D. and Blanke, D. (2007). *Inference and prediction in large dimensions*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester. MR2364006 (2008h:62010)

Bosq, D. and Onzon, E. (2012). Asymptotically efficient statistical predictors. *Journal de la Société Française de Statistique* 153 22–43.

Dedecker, J., Doukhan, P., Lang, G., León R., J. R., Louhichi, S. and Prieur, C. (2007). *Weak dependence: with examples and applications*. Lecture Notes in Statistics 190. Springer, New York. MR2338725 (2009a:62009)

Doukhan, P. (1994). *Mixing: Properties and examples*. Lecture Notes in Statistics 85. Springer-Verlag, New York. MR1312160 (96b:60090)

Drost, F. C., van den Akker, R. and Werker, B. J. M. (2008). Local asymptotic normality and efficient estimation for INAR(\( p \)) models. *J. Time Ser. Anal.* 29 783–801. MR2450896 (2009k:62210)

Garel, B. and Hallin, M. (1995). Local asymptotic normality of multivariate ARMA processes with a linear trend. *Ann. Inst. Statist. Math.* 47 551–579. MR1364260 (97b:62141)

Hájek, J. (1969/1970). A characterization of limiting distributions of regular estimates. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 14 323–330. MR0283911 (44 #91141)

Hájek, J. (1972). Local asymptotic minimax and admissibility in estimation. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971),* Vol. 1: Theory of statistics 175–194. Univ. California Press, Berkeley, Calif. MR0400513 (53 #94344)

Höpfner, R. (1988). Asymptotic inference for continuous-time Markov chains. *Probab. Theory Related Fields* 77 537–550. MR0933988 (89h:60144)
HÖPFNER, R., JACOD, J. and LADELLI, L. (1990). Local asymptotic normality and mixed normality for Markov statistical models. *Probab. Theory Related Fields* **86** 105–129. MR1061951 (91i:62035)

IBRAMIMOV, I. A. and HAS’MINSKII, R. Z. (1981). *Statistical estimation. Applications of Mathematics* **16**. Springer-Verlag, New York. Asymptotic theory, Translated from the Russian by Samuel Kotz. MR620321 (82g:62006)

JACOD, J. and SHIRYAEV, A. N. (1987). *Limit theorems for stochastic processes. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **288**. Springer-Verlag, Berlin. MR959133 (89k:60044)

KOUL, H. L. and SCHICK, A. (1997). Efficient estimation in nonlinear autoregressive time-series models. *Bernoulli* **3** 247–277. MR1468305 (99a:62132)

KREISS, J.-P. (1987). On adaptive estimation in autoregressive models when there are nuisance functions. *Statist. Decisions* **5** 59–76. MR886878 (88e:62213)

KUTOYANTS, Y. A. (2004). *Statistical inference for ergodic diffusion processes. Springer Series in Statistics*. Springer-Verlag London Ltd., London. MR2141581 (54 #3896)

LE CAM, L. (1986). *Asymptotic methods in statistical decision theory. Springer Series in Statistics*. Springer-Verlag, New York. MR856411 (88a:62004)

LEHMANN, E. L. and CASELLA, G. (1998). *Theory of point estimation*, second ed. *Springer Texts in Statistics*. Springer-Verlag, New York. MR1639875 (99g:62025)

MIYATA, Y. (2001). The lower bound for MSE in statistical prediction theory. *J. Japan Statist. Soc.* **31** 111–127.

NAYAK, T. K. (2002). Rao-Cramer type inequalities for mean squared error of prediction. *Amer. Statist.* **56** 102–106. MR1945869 (2003m:62225)

ONZON, E. (2011). Multivariate Cramér-Rao inequality for prediction and efficient predictors. *Statistics & Probability Letters* **81** 429–437.

PFANZAGL, J. (1994). *Parametric statistical theory. de Gruyter Textbook*. Walter de Gruyter & Co., Berlin. With the assistance of R. Hamböker. MR1291393 (96c:62001)

SWENSEN, A. R. (1985). The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. *J. Multivariate Anal.* **16** 54–70. MR778489 (86i:62036)

TANIGUCHI, M. and KAKIZAWA, Y. (2000). *Asymptotic theory of statistical inference for time series. Springer Series in Statistics*. Springer-Verlag, New York. MR1785484 (2002a:62131)

VAN DER VAART, A. W. (1998). *Asymptotic statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge University Press, Cambridge. MR1652247 (2000c:62003)

VERETENNIKOV, A. Y. (1987). Bounds for the Mixing Rate in the Theory of Stochastic Equations. *Theory of Probability and its Applications* **32** 273–281.

YATRACOS, Y. G. (1992). On prediction and mean squared error. *Canad. J. Statist.* **20** 187–200. MR1183080 (93k:62020)

Bâtiment Polytech Lyon, 15 Boulevard André Latarget, 69100 Villeurbanne, France

E-mail: emmanuel.onzon@univ-lyon1.fr