Acta Math. Univ. Comenianae
Vol. LXVI, 1(1997), pp. 71–81

PREDICTIONS IN NONLINEAR REGRESSION MODELS

F. ŠTULAJTER

Abstract. Different predictors and their approximators in nonlinear prediction regression models are studied. The minimal value of the mean squared error (MSE) is derived. Some approximate formulae for the MSE of ordinary and weighted least squares predictors are given.

1. Introduction

Theory of prediction is usually considered in linear regression models and there exist a lot of literature dealing with these problems, a.g. Journel (1977), Journel and Huijbregts (1978), Harville (1990) and others.

An open problem in this topic of prediction is the problem of finding an exact expression for the mean squared error of prediction in the case when the covariance characteristics of the prediction model are unknown and must be estimated from the data. Some approximate formulae for this case are given in Harville (1990), Harville and Jeske (1992), Bhansali (1981) and others.

In this time there is a lack of knowledge on methods of prediction and their properties in nonlinear regression models. The main reason for this fact probably is that the theory of parameter estimation in nonlinear regression models was developed mainly for uncorrelated errors. Some results of the author and others, see Štulajter (1992), Gallant (1987), Pázman (1993) show that the usual least squares estimators of regression parameters behave well (under some conditions similar to those known for uncorrelated errors) also for correlated errors, at least asymptotically. This is the reason for using these estimates also in problems of prediction.

These problems will be studied under the assumption that the covariance parameters of the prediction model are known. Althought there are some results about estimation of covariance parameters in nonlinear regression models, see Štulajter (1994), the problems of prediction with estimated covariance parameters are not solved even in linear regression models and thus we'll not study these problems in nonlinear regression models.

Received June 1, 1995.
1980 *Mathematics Subject Classification* (1991 *Revision*). Primary 62M20; Secondary 62M10.
Key words and phrases. Prediction regression model, mean squared error of prediction, ordinary and weighted least squares predictors.

The main effort of this article is to give (approximate) expressions for a class of predictors of a special type in a nonlinear regression model.

Let us consider the following prediction problem. Let us assume that observed random variables $X(t);\ t = 1, 2, \dots, n$ follow the model

$$X(t) = f(t, \theta) + \varepsilon(t); \qquad t = 1, 2, \dots, n \tag{1}$$

where f is a known function, θ is an unknown regression parameter, $\theta \in \Theta$ where Θ is an open subset of E^k and $\varepsilon(\cdot)$ are (random) errors with $E[\varepsilon(t)] = 0$, $E[\varepsilon(s)\varepsilon(t)] = R(s, t);\ s, t = 1, 2, \dots, n$, where $R(\cdot, \cdot)$ is a known covariance function of $X = \{X(t); t = 1, 2, \dots\}$. Let U be an unobservable random variable with $E_\theta[U] = g(\theta)$; $\theta \in \Theta$, where g is a known function and let $r(t) = \mathrm{Cov}\,(U; X(t));\ t = 1, 2, \dots, n$ be also known.

The problem is, on the base of $X(1), \dots, X(n)$ to predict the unknown random variable U and to find the mean squared error of prediction $E_\theta[U - \tilde{U}]^2$ of some predictor $\tilde{U}$.

Let us denote by $\mathbf{X}$ the random vector $\mathbf{X} = (X(1), \dots, X(n))'$, let $\mathbf{f}(\theta) = (f(1, \theta), \dots, f(n, \theta))';\ \theta \in \Theta$, $\varepsilon = (\varepsilon(1), \dots, \varepsilon(n))'$ and $\mathbf{r} = (r(1), \dots, r(n))'$. Then the **prediction regression model** (PRM) can be written as

$$\begin{aligned} &\mathbf{X} = \mathbf{f}(\theta) + \varepsilon; \quad E[\varepsilon] = \mathbf{O}, \quad E[\varepsilon\varepsilon'] = \Sigma, \\ &U = g(\theta) + \varkappa; \quad \mathrm{Cov}\,(\varepsilon; \varkappa) = \mathrm{Cov}\,(\mathbf{X}; U) = \mathbf{r}; \quad \theta \in \Theta \end{aligned} \tag{2}$$

where $\Sigma_{s,t} = R(s, t);\ s, t = 1, 2, \dots, n$. Let us assume that Σ is a positive definite matrix.

Definition. We call the prediction regression model linear if

$$\mathbf{f}(\theta) = \mathbf{F}\theta \quad \text{and} \quad g(\theta) = \mathbf{G}\theta; \quad \theta \in \Theta \tag{3}$$

where $\mathbf{F}$ is an $n \times k$ and $\mathbf{G}$ an $1 \times k$ known matrix. If the PRM is not linear we call it nonlinear. We call the PRM Gaussian if $(\mathbf{X}', U)'$ is a Gaussian random vector.

In the sequel we'll investigate properties of predictors $\tilde{U}$ which are given by

$$\tilde{U}(\tilde{\theta}) = g(\tilde{\theta}) + \mathbf{r}'\Sigma^{-1}(\mathbf{X} - \mathbf{f}(\tilde{\theta})) \tag{4}$$

where $\tilde{\theta}$ are some estimators of θ in the prediction regression model (2).

Example. Let us consider a time series $X = \{X(t);\ t = 1, 2, \dots\}$ with a mean value function

$$m_\theta(t) = \sum_{i=1}^{l} \alpha_i \cos(t\gamma_i + \beta_i); \quad t = 1, 2, \dots$$

depending on the regression parameter $\theta = (\alpha_1, \dots, \alpha_l, \beta_1, \dots, \beta_l, \gamma_1, \dots, \gamma_l)'$ and with some known covariance function $R(\cdot, \cdot)$. Let $\mathbf{X} = (X(1), \dots, X(n))'$ be a finite observation of X and let $U = X(n+1)$. The problem is to predict the unknown future value of the time series X using the observation $\mathbf{X}$. Then

$$f(t, \theta) = m_\theta(t); \quad t = 1, 2, \dots, n, \quad g(\theta) = m_\theta(n+1)$$

and, if X is covariance stationary with a covariance function R, $\Sigma_{s,t} = R(s-t)$ and $\mathbf{r} = (R(n), \dots, R(1))'$.

The motivation for using the predictors $\tilde{U}$ given by (4) is in the next section.

2. Predictors in Prediction Regression Models

Let us consider the prediction regression model (2) with $\mathbf{r}$ and Σ known, Σ a positive definite matrix. The dependence of $\mathbf{f}$ and g on θ can be linear or nonlinear. Let $\tilde{U}$ be any, generally nonlinear, predictor of U . Then the mean squared error of predictor is given by

$$\begin{aligned} \text{(5)} \qquad \mathrm{MSE}_\theta[\tilde{U}] &= E_\theta[U - \tilde{U}]^2 \\ &= D[U] + D_\theta[\tilde{U}] + (E_\theta[U] - E_\theta[\tilde{U}])^2 - 2\mathrm{Cov}\,(U; \tilde{U}); \quad \theta \in \Theta. \end{aligned}$$

The $\mathrm{MSE}_\theta[\tilde{U}]$ of a predictor $\tilde{U}$ does not depend on θ, the mean value parameter, iff the variance of $\tilde{U}$ does not depend on θ and $\tilde{U}$ is an unbiased predictor that means $\tilde{U}$ is such that

$$E_\theta[\tilde{U}] = g(\theta) \text{ for all } \theta \in \Theta.$$

Now, let us consider a linear predictor $\tilde{U}$ of U of the form

$$\tilde{U}(\mathbf{X}) = \mathbf{a}'\mathbf{X} + b, \text{ where } \mathbf{a} \in E^n, \ b \in E^1.$$

Then we have

$$E_\theta[\tilde{U}] = \mathbf{a}'\mathbf{f}(\theta) + b; \quad \theta \in \Theta$$

and the unbiasedness condition for $\tilde{U}$ is given by

$$\text{(6)} \qquad \mathbf{a}'\mathbf{f}(\theta) + b = g(\theta) \text{ for all } \theta \in \Theta.$$

For a linear prediction regression model, with $\mathbf{f}$ and g given by (3), the unbiasedness condition (6) can be written as

$$\mathbf{a}'\mathbf{F}\theta + b = \mathbf{G}\theta; \quad \theta \in \Theta,$$

or equivalently as

$$\mathbf{F}'\mathbf{a} = \mathbf{G}', \quad b = 0$$

and this condition can be always fulfilled if $\mathbf{F}$ has rank k.

For given nonlinear regression functions $\mathbf{f}$ and g the condition (6) need not be fulfilled (for all $\theta \in \Theta$) and thus for a nonlinear prediction regression model unbiased estimator in general need not to exist.

Now, let θ_0 be a fixed value of the parameter θ. Then there exists a class U_0 of linear locally (at θ_0) unbiased predictors of U given by

$$U_0 = \left\{\tilde{U} = \mathbf{a}'\mathbf{X} + b;\ b = g(\theta_0) - \mathbf{a}'\mathbf{f}(\theta_0);\ \mathbf{a} \in E^n\right\}.$$

For a predictor $\tilde{U} \in U_0$ we have:

$$\mathrm{MSE}_{\theta_0}[\tilde{U}] = D[U] + D[\tilde{U}] - 2\ \mathrm{Cov}\,(U;\tilde{U}) = D[U] + \mathbf{a}'\Sigma\mathbf{a} - 2\mathbf{a}'\mathbf{r}$$

and this $\mathrm{MSE}_{\theta_0}[\tilde{U}]$ is minimized, with respect to $\mathbf{a}$, by setting $\mathbf{a} = \mathbf{a}^* = \Sigma^{-1}\mathbf{r}$. To this $\mathbf{a}^*$ we have the corresponding predictor $U_0^* = (\mathbf{a}^*)'\mathbf{X} + g(\theta_0) - (\mathbf{a}^*)'\mathbf{f}(\theta_0)$.

Definition. The predictor U_0^* given by

$$U_0^* = g(\theta_0) + \mathbf{r}'\Sigma^{-1}(\mathbf{X} - \mathbf{f}(\theta_0)) \tag{7}$$

will be called **the best linear locally** (at θ_0) **unbiased predictor** (BLLUP) of U for which we have

$$\mathrm{MSE}_{\theta_0}[U_0^*] = D[U] - \mathbf{r}'\Sigma^{-1}\mathbf{r}. \tag{8}$$

Remark. It is well known that if $(\mathbf{X}', U)'$ is a Gaussian random vector then the **linear** predictor U_0^* is the best locally (at θ_0) unbiased predictor minimizing the $\mathrm{MSE}_{\theta_0}[\tilde{U}]$ among all, not only linear, locally (at θ_0) unbiased predictors $\tilde{U}$.

This predictor has the disadvantage that it depends on the value θ_0 which is, usually, unknown in practice and thus U_0^* can not be used in real situations, not even in linear prediction regression models.

But, it is well known, see Journel (1977) or Harville (1990), that for a linear prediction regression model there exists the (uniformly) **best linear unbiased predictor** (BLUP) U^* given by

$$U^* = g(\theta^*) + \mathbf{r}'\Sigma^{-1}(\mathbf{X} - \mathbf{f}(\theta^*) \tag{9}$$

where $\theta^* = (\mathbf{F}'\Sigma^{-1}\mathbf{F})^{-1}\mathbf{F}'\Sigma^{-1}\mathbf{X}$ is the best linear unbiased estimator (BLUE) of θ, for which

$$\mathrm{MSE}_\theta[U^*] = D[U] - \mathbf{r}'\Sigma^{-1}\mathbf{r} + ||\mathbf{G}' - \mathbf{F}'\Sigma^{-1}\mathbf{r}||^2_{\Sigma_{\theta^*}}$$

does not depend on θ. Here $\Sigma_{\theta^*} = (\mathbf{F}'\Sigma^{-1}\mathbf{F})^{-1}$ denotes the covariance matrix of θ^* and the norm is defined by $||\mathbf{a}||^2_{\Sigma_{\theta^*}} = \mathbf{a}'\Sigma_{\theta^*}\mathbf{a}$; $\mathbf{a} \in E^k$. We see that in linear prediction models we get the BLUP from the BLLUP by substituting the BLUE θ^* for the unknown parameter θ_0.

This result motivates us to use the predictors $\tilde{U}$ given by (4) also in nonlinear prediction regression models. We'll investigate their properties in the next section.

3. Properties of Predictors in a PRM

Let us consider the (nonlinear) prediction regression model given by (2) and let $\tilde{U}$ be the predictor given by (4) with a corresponding mean squared error given by (5) . Since in nonliner regression models good estimators $\tilde{\theta}$ of θ are nonlinear, predictors $\tilde{U}$ are usually also nonlinear. As a rule, we do not have an explicit expression for nonlinear estimators and their moments because they are computed iteratively and thus we can not compute directly the mean squared error of a nonlinear predictor. The problem is that in (5) the exact expressions for expectation, variance and covariance of an predictor $\tilde{U}$ depend on θ (which is unknown) and are unknown. We'll give some approximations to these expressions based on corresponding approximations for the functions $\mathbf{f}$ and g and for $\tilde{\theta}$ which enable us to give approximate expressions for the $\mathrm{MSE}_\theta[\tilde{U}]$.

We'll turn our attention to the following two predictors.

Definition. Let $\hat{\theta}$ be the ordinary least squares estimator (LSE) of θ minimizing $||\mathbf{X} - \mathbf{f}(\theta)||^2_I$. Then the predictor $\hat{U}$ given by $\hat{U} = \tilde{U}(\hat{\theta})$ will be called the **ordinary least squares predictor** (OLSEP) of U. Let θ^* be the weighted LSE of θ minimizing $||\mathbf{X} - \mathbf{f}(\theta)||^2_{\Sigma^{-1}}$. Then the EP U^* given by $U^* = \tilde{U}(\theta^*)$ will be called the **weighted least squares predictor** (WELSEP) of U.

In the sequel we'll consider the PRM given by (2) with such regression functions $\mathbf{f}$ and g that for every $\theta \in \Theta$ there exist $\frac{\partial f(t,\theta)}{\partial \theta}$; $t = 1, 2, \ldots, k$ and $\frac{\partial g(\theta)}{\partial \theta}$.

Let $\mathbf{F}(\theta)$ denotes the $n \times k$ matrix with $\mathbf{F}(\theta)_{ti} = \frac{\partial f(t,\theta)}{\partial \theta_i}$ and $\mathbf{G}(\theta)$ the $1 \times k$ matrix with $\mathbf{G}(\theta)_{1i} = \frac{\partial g(\theta)}{\partial \theta_i}$; $t = 1, 2, \ldots n$; $i = 1, 2, \ldots k$. Then we can, for every fixed θ_0, approximate any predictor $\tilde{U}$ given by (4) by

$$\tilde{U}_L = g(\theta_0) + \mathbf{G}_0(\tilde{\theta} - \theta_0) + \mathbf{r}'\Sigma^{-1}(\mathbf{X} - \mathbf{f}(\theta_0) - \mathbf{F}_0(\theta - \theta_0))$$

where $\mathbf{G}_0 = \mathbf{G}(\theta_0)$ and $\mathbf{F}_0 = \mathbf{G}(\theta_0)$. Using the linear approximation $\tilde{\theta}_L$ given by

$$\tilde{\theta}_L = \theta_0 + \tilde{\mathbf{A}}_0\varepsilon$$

(where the $k \times n$ matrix $\tilde{\mathbf{A}}_0$ depends on a type of the estimator $\tilde{\theta}$ and on θ_0) for the estimator $\tilde{\theta}$ we get the linear approximation $\tilde{U}_{L,L}$ of $\tilde{U}$ given by

$$\tilde{U}_{L,L} = g(\theta_0) + \mathbf{G}_0\tilde{\mathbf{A}}_0\varepsilon + \mathbf{r}'\Sigma^{-1}\tilde{\mathbf{M}}_0\varepsilon \tag{10}$$

where $\tilde{\mathbf{M}}_0 = I - \tilde{\mathbf{P}}_0$ with $\tilde{\mathbf{P}}_0 = \mathbf{F}_0\tilde{\mathbf{A}}_0$.

$\tilde{U}_{L,L}$, expressed in terms of the error vector ε, can be regarded as a local (at $\theta = \theta_0$) linear approximation of the predictor $\tilde{U}$ and can be used for computing an approximate formula for the $\mathrm{MSE}_{\theta_0}[\tilde{U}]$ (for any $\theta_0 \in \Theta$). We see from (10) that $\tilde{U}_{L,L}$ is locally (at θ_0) unbiased and it is easy to compute, using (5) and some matrix algebra, that

$$\mathrm{MSE}_{\theta_0}[\tilde{U}_{L,L}] = D[U] - \mathbf{r}'\Sigma^{-1}\mathbf{r} + ||\mathbf{G}_0' - \mathbf{F}_0'\Sigma^{-1}\mathbf{r}||^2_{\tilde{\mathbf{A}}_0\Sigma\tilde{\mathbf{A}}_0'} \tag{11}$$

Now, let us consider the OLSEP $\hat{U}$ and the WELSEP U^*. Then $\hat{U}_{L,L}$, the linear approximation of $\hat{U}$, is given by (10) with $\tilde{\mathbf{A}}_0 = \hat{\mathbf{A}}_0 = (\mathbf{F}_0'\mathbf{F}_0)^{-1}\mathbf{F}_0'$ (if we suppose that $\mathbf{F}_0$ has rank k) and

$$\text{(12)} \qquad \mathrm{MSE}_{\theta_0}[\hat{U}_{L,L}] = D[U] - \mathbf{r}'\Sigma^{-1}\mathbf{r} + \|\mathbf{G}_0' - \mathbf{F}_0'\Sigma^{-1}\mathbf{r}\|^2_{\Sigma_{\hat{\theta}_L}}$$

where $\Sigma_{\hat{\theta}_L} = (\mathbf{F}_0'\mathbf{F}_0)^{-1}\mathbf{F}_0'\Sigma\mathbf{F}_0(\mathbf{F}_0'\mathbf{F}_0)^{-1}$ is the covariance matrix of the linear approximation $\hat{\theta}_L = (\mathbf{F}_0'\mathbf{F}_0)^{-1}\mathbf{F}_0'\varepsilon$ of the LSE $\hat{\theta}$.

The WELSEP U^* of U can be approximated by $U^*_{L,L}$ given by

$$U^*_{L,L} = g(\theta_0) + \mathbf{G}_0\mathbf{A}_0^*\varepsilon + \mathbf{r}'\Sigma^{-1}\mathbf{M}_0^*\varepsilon$$

where

$$\mathbf{A}_0^* = (\mathbf{F}_0'\Sigma^{-1}\mathbf{F}_0)^{-1}\mathbf{F}_0'\Sigma^{-1}\varepsilon \quad \text{and} \quad \mathbf{M}_0^* = I - \mathbf{P}_0^* = I - \mathbf{F}_0(\mathbf{F}_0'\Sigma^{-1}\mathbf{F}_0)^{-1}\mathbf{F}_0'\Sigma^{-1}$$

with

$$\text{(13)} \qquad \mathrm{MSE}_{\theta_0}[U^*_{L,L}] = D[U] - \mathbf{r}'\Sigma^{-1}\mathbf{r} + \|\mathbf{G}_0' - \mathbf{F}_0'\Sigma^{-1}\mathbf{r}\|^2_{\Sigma_{\theta^*_L}}$$

where $\Sigma_{\theta^*_L} = (\mathbf{F}_0'\Sigma^{-1}\mathbf{F})^{-1}$ is the covariance matrix of the linear approximation $\theta^*_L = (\mathbf{F}_0'\Sigma^{-1}\mathbf{F}_0)^{-1}\mathbf{F}_0'\Sigma^{-1}\varepsilon$ of the weighted LSE θ^*. From the results mentioned in Section 2 (for linear prediction regression models) we have:

Lemma 1. *For the linear approximations $\hat{U}_{L,L}$ and $U^*_{L,L}$ of the OLSEP and of the WELSEP, respectively, we have the inequalities:*

$$D[U] - \mathbf{r}'\Sigma^{-1}\mathbf{r} \le \mathrm{MSE}_\theta[U^*_{L,L}] \le \mathrm{MSE}_\theta[\hat{U}_{L,L}] \quad \textit{for every } \theta \in \Theta.$$

Remark. If the model (2) is a linear prediction regression model then $\hat{U}$ and U^* are linear predictors and

$$\mathrm{MSE}_\theta[\hat{U}] = \mathrm{MSE}_\theta[\hat{U}_{L,L}] \quad \text{and} \quad \mathrm{MSE}_\theta[U^*] = \mathrm{MSE}_\theta[U^*_{L,L}] \ \text{ for every } \theta \in \Theta.$$

The preceding lemma does not tell us how good the approximations $\hat{U}_{L,L}$ and $U^*_{L,L}$ of $\hat{U}$ and U^*, respectively, are. If the PRM is nonlinear, the mean squared errors of these approximations can be substantially different from those of predictors $\hat{U}$ and U^*. In such models we have to use more accurate approximations.

Let us assume that for every $\theta_0 \in \Theta$ there exist a $k \times k$ Hessian matrix $\mathbf{H}_g$ with the elements $(\mathbf{H}_g)_{ij} = \frac{\partial^2 g(\theta)}{\partial\theta_i\partial\theta_j}|_{\theta=\theta_0}$ and for every $t = 1, 2, \dots, n$ the Hessian matrices $\mathbf{H}(t)$ with the elements $(\mathbf{H}(t))_{ij} = \frac{\partial^2 f(t,\theta)}{\partial\theta_i\,\partial\theta_j}|_{\theta=\theta_0}$; $i, j = 1, 2, \dots, k$.[1]

[1] For simplicity of notation we'll not denote the dependence of the corresponding matrices on θ_0 in the sequel.

Then we can write

$$g(\tilde{\theta}) = g(\theta_0) + \mathbf{G}(\tilde{\theta} - \theta_0) + \frac{1}{2}(\tilde{\theta} - \theta_{0'})\mathbf{H}_g(\tilde{\theta} - \theta_0)$$

$$\mathbf{f}(\tilde{\theta}) = \mathbf{f}(\theta_0) + \mathbf{F}(\tilde{\theta} - \theta_0) + \frac{1}{2}\left\langle(\tilde{\theta} - \theta_{0'})\mathbf{H}(t)(\tilde{\theta} - \theta_0)\right\rangle$$

where $\langle\alpha'\mathbf{H}(t)\alpha\rangle$; $\alpha \in E^k$ denotes the $n \times 1$ vector with components $\alpha'\mathbf{H}(t)\alpha$; $t = 1, 2, \dots, n$. Using a linear approximation $\tilde{\theta}_L = \theta_0 + \tilde{\mathbf{A}}\varepsilon$ for $\tilde{\theta}$ we get a quadratic-linear approximation $\tilde{U}_{Q,L}$ of $\tilde{U}$ given by

$$\tilde{U}_{Q,L} = g(\theta_0) + \mathbf{G}\tilde{A}\varepsilon + \frac{1}{2}\varepsilon'\tilde{\mathbf{A}}'\mathbf{H}_g\tilde{\mathbf{A}}\varepsilon$$

$$+ \mathbf{r}'\Sigma^{-1}\left(\mathbf{X} - \mathbf{f}(\theta_0) - \mathbf{F}\tilde{A}\varepsilon\frac{1}{2}\left\langle\varepsilon'\tilde{\mathbf{A}}'\mathbf{H}(t)\tilde{\mathbf{A}}\varepsilon\right\rangle\right) \quad \text{or}$$

$$\tilde{U}_{Q,L} = \tilde{U}_{L,L} + \tilde{Q}_1 \tag{14}$$

where $\tilde{Q}_1(\varepsilon) = \frac{1}{2}[\varepsilon'\tilde{\mathbf{A}}'\mathbf{H}_g\tilde{\mathbf{A}}\varepsilon - \mathbf{r}'\Sigma^{-1}\langle\varepsilon'\tilde{\mathbf{A}}'\mathbf{H}(t)\tilde{\mathbf{A}}\varepsilon\rangle]$ is a linear combination of quadratic (in ε) forms.

Using the expression $E[\varepsilon'\mathbf{B}\varepsilon] = \mathrm{tr}\,(\mathbf{B}\Sigma)$ which holds for every $n \times n$ symmetric matrix $\mathbf{B}$ and for every $n \times 1$ random vector ε with mean value zero and with a covariance matrix Σ we get

$$E_{\theta_0}[\tilde{U}_{Q,L}] = g(\theta_0) + \frac{1}{2}\left(\mathrm{tr}\,(\tilde{\mathbf{A}}'\mathbf{H}_g\tilde{\mathbf{A}}\Sigma) - \mathbf{r}'\Sigma^{-1}\left\langle\mathrm{tr}\,\tilde{\mathbf{A}}'\mathbf{H}(t)\tilde{\mathbf{A}}\Sigma)\right\rangle\right) \tag{15}$$

We see that $\tilde{U}_{Q,L}$ is no more (locally) unbiased approximation of $\tilde{U}$, its bias, as it follows from (15) depends on θ_0, on the approximation matrix $\tilde{\mathbf{A}}$ and on Hessian matrices $\mathbf{H}_g$ and $\mathbf{H}(t)$, $t = 1, 2, \dots, n$. Of course, it depends also on covariance parameters $\mathbf{r}$ and Σ of the PRM (2).

The $\mathrm{MSE}_{\theta_0}[\tilde{U}_{Q,L}]$ can be computed using (5) and (14).

We get, after some computation

$$\mathrm{MSE}_{\theta_0}[\tilde{U}_{Q,L}] = \mathrm{MSE}_{\theta_0}[\tilde{U}_{L,L}] + E_{\theta_0}[\tilde{Q}_1^2] + 2\mathrm{Cov}\,(\tilde{U}_{L,L};\tilde{Q}_1) - 2\mathrm{Cov}\,(U;\tilde{Q}_1).$$

From this expression we get the following lemma.

Lemma 2. *Let the random errors ε and $\varkappa$ in the PRM (2) fulfills the equalities $E[\varkappa\varepsilon'\mathbf{B}\varepsilon] = E[\varepsilon_t\varepsilon'\mathbf{B}\varepsilon] = 0$ for every $n \times n$ symmetric matrix $\mathbf{B}$ and every $t = 1, 2, \dots, n$. Then*

$$MSE_\theta[\tilde{U}_{Q,L}] = MSE_\theta[\tilde{U}_{L,L}] + E_\theta[\tilde{Q}_1^2]; \quad \theta \in \Theta. \tag{16}$$

Remark. The conditions of Lemma 2 are fulfilled if the PRM (2) is Gaussian.

In this case we can use the equality

$$E[\varepsilon'\mathbf{B}\varepsilon\varepsilon'\mathbf{C}\varepsilon] = 2\mathrm{tr}\,(\mathbf{B}\Sigma C\Sigma) + \mathrm{tr}\,(\mathbf{B}\Sigma)\mathrm{tr}\,(\mathbf{C}\Sigma)$$

which holds, see Kubáček (1988), for any $n \times n$ symmetric matrices $\mathbf{B}$ and $\mathbf{C}$. Using this equality we get, after some computation the following theorem.

Theorem 1. *Let us suppose that the PRM (2) is Gaussian. Then*

$$\begin{aligned}\mathrm{MSE}_\theta[\tilde{U}_{Q,L}] &= \mathrm{MSE}_\theta\Big[\tilde{U}_{L,L}] + \frac{1}{4}[2tr(\mathbf{B}\Sigma B\Sigma)\\ &\quad + (\mathrm{tr}\,(\mathbf{B}\Sigma))^2 + \mathbf{r}'\Sigma^{-1}\left\langle 2\mathrm{tr}\,(\mathbf{C}_s\Sigma\mathbf{C}_t\Sigma) + \mathrm{tr}\,(\mathbf{C}_s\Sigma)\mathrm{tr}\,(\mathbf{C}_t\Sigma)\right\rangle\Sigma^{-1}\mathbf{r}\\ &\quad - 2\mathbf{r}'\Sigma^{-1}\left\langle 2\mathrm{tr}\,(\mathbf{C}_t\Sigma\mathbf{B}\Sigma) + \mathrm{tr}\,(\mathbf{C}_t\Sigma)\mathrm{tr}\,(\mathbf{B}\Sigma)\right\rangle\Big]\\ &\geq \mathrm{MSE}_\theta[\tilde{U}_{L,L}]\end{aligned}$$

where $\mathbf{B} = \tilde{\mathbf{A}}'\mathbf{H}_g\tilde{\mathbf{A}}$, $\mathbf{C}_t = \tilde{\mathbf{A}}'\mathbf{H}(t)\tilde{\mathbf{A}}$ *and* $\langle a_{st}\rangle$ *denotes the* $n \times n$ *matrix with elements* a_{st}*;* $s, t = 1, 2, \ldots, n$.

Remark. Using the corresponding matrices $\hat{\mathbf{A}}$ and $\mathbf{A}^*$ for $\tilde{\mathbf{A}}$ in (17) we get the $\mathrm{MSE}_\theta[\hat{U}_{Q,L}]$ and $\mathrm{MSE}[U^*_{Q,L}]$ of the approximations $\hat{U}_{Q,L} = \tilde{U}_{Q,L}(\hat{\theta})$ and $U^*_{Q,L} = \tilde{U}_{Q,L}(\theta^*)$ of U.

Till now we have used only linear approximation $\tilde{\theta}_L = \theta_0 + \tilde{\mathbf{A}}\varepsilon$ for the nonlinear estimator $\tilde{\theta}$ of θ. But, as it was shown in Box (1971) and in a more convenient form in Štulajter (1992), it is possible to use for $\tilde{\theta}$ also a quadratic approximation $\tilde{\theta}_Q$ given by

$$\tilde{\theta}_Q(\varepsilon) = \theta_0 + \tilde{\mathbf{A}}\varepsilon + \tilde{\mathbf{Q}}(\varepsilon) \tag{18}$$

where $\tilde{\mathbf{Q}}$ is of the form

$$\tilde{\mathbf{Q}}(\varepsilon) = \tilde{\mathbf{D}}\left(\left\langle\varepsilon'\tilde{\mathbf{N}}_i\varepsilon\right\rangle - \frac{1}{2}\tilde{\mathbf{K}}\left\langle\varepsilon'\tilde{\mathbf{A}}'\mathbf{H}(t)\tilde{\mathbf{A}}\varepsilon\right\rangle\right) \tag{19}$$

where $\tilde{\mathbf{D}}$ is a $k \times k$ matrix, $\tilde{\mathbf{N}}_i$; $i = 1, 2, \ldots, k$ are $n \times n$ matrices and $\tilde{\mathbf{K}}$ is a $k \times n$ matrix.

Thus, using the linear approximation for regression functions $\mathbf{f}$ and g and the quadratic approximation $\tilde{\theta}_Q$ given by (18), we obtain the approximation $\tilde{U}_{L,Q}$ of $\tilde{U}$ given by

$$\tilde{U}_{L,Q} = \tilde{U}_{L,L} + \mathbf{q}'\tilde{\mathbf{Q}} \tag{20}$$

where $\mathbf{q}$ is a $k \times 1$ vector given by

$$\mathbf{q} = \mathbf{G}' - \mathbf{F}'\Sigma^{-1}\mathbf{r}. \tag{21}$$

For computing the $\mathrm{MSE}_\theta[\tilde{U}_{L,Q}]$ we can use again (5) and (20). We get after some computation

$$\mathrm{MSE}_\theta[\tilde{U}_{L,Q}] = \mathrm{MSE}_\theta[\tilde{U}_{L,L}] + E_\theta[(\mathbf{q}'\tilde{\mathbf{Q}})^2] + 2\mathrm{Cov}\,(\tilde{U}_{L,L};\mathbf{q}'\tilde{\mathbf{Q}}) - 2\mathrm{Cov}\,(U;\mathbf{q}'\tilde{\mathbf{Q}}).$$

Comparing $\tilde{U}_{L,Q}$ given by (20) with $\tilde{U}_{Q,L}$ given by (14) we see that both "quadratic" terms $\mathbf{q}'\tilde{\mathbf{Q}}(\varepsilon)$ and $\tilde{Q}_1(\varepsilon)$ are linear combinations of quadratic forms. Thus we have, as before, the following lemma and theorem.

Lemma 3. *Let the assumptions of the Lemma 2 are fulfilled. Then*

$$\mathrm{MSE}_\theta[\tilde{U}_{L,Q}] = \mathrm{MSE}_\theta[\tilde{U}_{L,L}] + E_\theta[(\mathbf{q}'\tilde{\mathbf{Q}})^2].$$

Theorem 2. *Let us suppose that the PRM (2) is Gaussian. Then*

$$\begin{aligned}\mathrm{MSE}_\theta[\tilde{U}_{L,Q}] &= \mathrm{MSE}_\theta[\tilde{U}_{L,L}] + \mathbf{q}'\tilde{\mathbf{D}}[\langle 2\mathrm{tr}\,(\mathbf{N}_i\Sigma\mathbf{N}_j\Sigma) + \mathrm{tr}\,(\mathbf{N}_i\Sigma)\mathrm{tr}\,(\mathbf{N}_j\Sigma)\rangle \\ &\quad - \tilde{\mathbf{K}}\,\langle 2\mathrm{tr}\,(\mathbf{C}_t\Sigma\mathbf{N}_i\Sigma) + \mathrm{tr}\,(\mathbf{C}_t\Sigma)\mathrm{tr}\,(\mathbf{N}_i\Sigma)\rangle \\ &\quad + \frac{1}{4}\tilde{\mathbf{K}}\,\langle 2\mathrm{tr}\,(\mathbf{C}_t\Sigma\mathbf{C}_s\Sigma) + \mathrm{tr}\,(\mathbf{C}_t\Sigma)\mathrm{tr}\,(\mathbf{C}_s\Sigma)\rangle\,\tilde{\mathbf{K}}']\tilde{\mathbf{D}}'\mathbf{q} \\ &\geq \mathrm{MSE}_\theta[\tilde{U}_{L,L}],\end{aligned}$$

where, as before, $\mathbf{C}_t = \tilde{\mathbf{A}}'\mathbf{H}(t)\tilde{\mathbf{A}}$.

Remark. We can write:

$$\begin{aligned}\mathrm{MSE}_\theta[\tilde{U}_{L,Q}] &= \mathrm{MSE}_\theta[\tilde{U}_{L,L}] + ||\mathbf{G}' - \mathbf{F}'\Sigma^{-1}\mathbf{r}||_V^2 \\ &= D[U] - \mathbf{r}'\Sigma^{-1}\mathbf{r} + ||\mathbf{G}' - \mathbf{F}'\Sigma^{-1}\mathbf{r}||^2_{\tilde{A}\Sigma\tilde{A}'+V}\end{aligned}$$

where the expression for $\mathbf{V} = E[\tilde{\mathbf{Q}}\tilde{Q}']$ follows from (22).

It is an open problem to compare the $\mathrm{MSE}_\theta[\tilde{U}_{Q,L}]$ and the $\mathrm{MSE}_\theta[\tilde{U}_{L,Q}]$.

It is shown in Štulajter (1992) that for $\hat{\theta}$, the OLSE of θ, its quadratic approximation $\hat{\theta}_Q$ is given by (18) and (19) with $\tilde{\mathbf{D}} = \hat{\mathbf{D}} = (\mathbf{F}'\mathbf{F})^{-1}$, $\tilde{\mathbf{K}} = \hat{\mathbf{K}} = \mathbf{F}'$ and $\tilde{\mathbf{N}}_i = \hat{\mathbf{N}}_i$ with $(\hat{\mathbf{N}}_i)_{jl} = \sum_{t=1}^n(\mathbf{H}(t)\hat{\mathbf{A}})_{ij}\hat{\mathbf{M}}_{tl}$; $j,l = 1,2,\ldots,n$, where, as before $\hat{\mathbf{A}} = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'$ and $\hat{\mathbf{M}} = \mathbf{I} - \hat{\mathbf{P}} = \mathbf{I} - \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'$. These matrices should be also used in (22) for computing the $\mathrm{MSE}_\theta[\hat{U}_{L,Q}]$.

Similar expressions can be given also for matrices defining the quadratic approximation $U^*_{L,Q}$ of the predictor U^*: $\mathbf{A}^* = \mathbf{A}^*_0$ and $\mathbf{M}^* = \mathbf{M}^*_0$, where $\mathbf{M}^*_0$ and $\mathbf{A}^*_0$ were already defined for $U^*_{L,L}$, $\mathbf{D}^* = (\mathbf{F}'\Sigma^{-1}\mathbf{F})^{-1}$, $\mathbf{K}^* = \mathbf{F}'\Sigma^{-1}$ and $(\mathbf{N}^*_i)_{jl} = \sum_{t=1}^n(\mathbf{H}(t)\mathbf{A}^*)_{ij}(\Sigma^{-1}\mathbf{M}^*)_{tl}$; $j,l = 1,2,\ldots,n$.

A comparison of $\hat{U}_{Q,L}$, $U^*_{Q.L}$, $\hat{U}_{L,Q}$ and $U^*_{L,Q}$ with respect to their mean squared errors is still an open problem.

Remark. It is possible to use quadratic approximations for prediction regression functions $\mathbf{f}$ and g and also for an estimator $\tilde{\theta}$ of regression parameter θ. Doing this we get an approximation $\tilde{U}_{Q,Q}$ of $\tilde{U}$ containing terms $\varepsilon'\tilde{\mathbf{A}}'\mathbf{J}\mathbf{Q}(\varepsilon)$ and $\mathbf{Q}(\varepsilon)'\mathbf{J}Q(\varepsilon)$ with $\mathbf{J}$ a symmetric $k \times k$ matrix, for which we have no explicit formulae for their variances. Thus we have no explicit expression for the $\mathrm{MSE}_\theta[\tilde{U}_{Q,Q}]$.

4. Conclusions

In the preceding parts of the article different approximations for empirical predictors $\tilde{U}$ of U were suggested and their mean squared errors were derived. As we've already mentioned a comparison of these approximations and their mean squared errors is difficult, since they depend on the model functions, on the true value of regression parameter and on covariance characteristics of the prediction regression model. Also a comparison of their mean squared errors with the $\mathrm{MSE}_\theta[\tilde{U}]$ is an open problem.

One possibility to clear these problems is to make simulation studies for some particular regression models (which are also often used in practical applications). These simulation studies can be used for a comparison of ordinary and weighted least squares predictors and their approximations. Such a comparison can be found in Štulajter and Stano. The open problem is also the problem of estimation the $\mathrm{MSE}_\theta[\tilde{U}]$, as it is, for a linear prediction regression model, studied in Harville (1985) and (1990).

The main problem, from the point of view of practical applications, is the problem of prediction in (nonlinear) prediction regression models with unknown covariance characterics which should also be estimated from the observations. In this connection the paper of Štulajter (1994) can serve as a base for further investigations. Another possible approach to this problem is to use a parametric regression model also for covariance characteristics of a prediction regression model. Some results on estimation of parameters in such models are given in Gumpertz and Pantulla (1992).

References

Box M. J., *Bias in nonlinear estimation*, Jour. Roy. Stat. Soc. **B 33** (1971), 171–201.

Bhansali R. J., *Effect of not knowing the order of autoregression on the mean squared error of prediction*, J. Amer. Stat. Assoc. **78** (1977), 588–597.

Gumpertz M. L. and Pantulla S. G., *Nonlinear regression with variance components*, J. Amer. Stat. Assoc. **87** (1992), 201–209.

Gallant A. R., *Nonlinear statistical modeling*, Wiley, New York, 1987.

Harville D. A., *Maximum likelihood approaches to variance components estimation and to related problems*, J. Amer. Stat. Assoc. **57** (1977), 320–338.

______, *Decomposition of prediction error*, J. Amer. Stat. Assoc. **80** (1985), 132–138.

______, *BLUP (best linear unbiased predictor) and beyond*, Advances in stat. methods for genetic improvement of livestock, Springer Verlag, New York, 1990, pp. 239–276.

Haville D. A. and Jeske D. R., *Mean squared error of estimation or prediction under a general linear model*, J. Amer. Stat. Assoc. **87** (1992), 724–731.

Journel A., *Kriging in the terms of predictions*, J. Inter. Assoc. Math. Geol. **9** (1977), 563–586.

Journel A. and Huijbregts C., *Mining geostatistics*, New York, 1978.

Kubáček L., *Foundations of estimation theory*, Amsterdam, Elsevier, 1988.

Pázman A., *Nonlinear statistical models*, Kluwer, Dodrecht, 1993.

Rattkowsky D. A., *Nonlinear regression modelling*, Marcel Decker, New York, 1983.

Štulajter F., *Consistency of linear and quadratic least squares estimators in regression models with covariance stationary errors*, Appl. Math. **36** (1991), 149–155.

______, *Mean squared error matrix of an approximate least squares estimator in a nonlinear regression model with correlated errors*, Acta Math. Univ. Comen. **LXI(2)** (1992), 251–261.

______, *Some aspects of kriging*, Transac. of the 11-th Prague conference, Prague, 397–402.

———, *On estimation of covariance function of stationary errors in a nonlinear regression model*, Acta Math. Univ. Comen. **LXIII(1)** (1994), 107–116.

Štulajter F. and Stano S., *A comparisom of some predictors in nonlinear regression models*, Submitted for publication.

Toyooka Y., *Prediction error in a linear model with estimated parameters*, Biometrica **69** (1982), 453–459.

Zimmerman D. L. and Cressie N., *Mean squared error in spatial linear models with estimated covariance parameters*, Ann. Inst. Stat. Math. **44** (1992), 27–43.

F. Štulajter, Department of Probability and Statistcs, Faculty of Mathematics and Physics, Comenius University, 842 15 Bratislava, Slovakia