Available online at www.sciencedirect.com
ScienceDirect
Statistics & Probability Letters 77 (2007) 811–816
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

# Difference-based estimation for error variances in repeated measurement regression models

Qinfeng Xu[a], Jinhong You[b,*]

[a]Department of Statistics, Fudan University, Shanghai 200433, PR China
[b]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7400, USA

Received 30 April 2006; received in revised form 2 October 2006; accepted 26 November 2006
Available online 1 February 2007

**Abstract**

Consider a repeated measurement regression model $y_{ij} = g(x_i) + \varepsilon_{ij}$ where $i = 1, \ldots, n$, $j = 1, \ldots, m$, $y_{ij}$'s are responses, $g(\cdot)$ is an unknown function, $x_i$'s are design points, $\varepsilon_{ij}$'s are random errors with a one-way error component structure, i.e. $\varepsilon_{ij} = \mu_i + v_{ij}$, $\mu_i$ and $v_{ij}$'s are i.i.d random variables with mean zero, variance $\sigma_\mu^2$ and $\sigma_v^2$, respectively. This paper focuses on estimating $\sigma_\mu^2$ and $\sigma_v^2$. It is well known that although the residual-based estimator of $\sigma_v^2$ works very well the residual-based estimator of $\sigma_\mu^2$ works poorly, especially when the sample size is small. We here propose a difference-based estimation and show the resulted estimator of $\sigma_\mu^2$ performs much better than the residual-based one. In addition, we show the difference-based estimator of $\sigma_v^2$ is equal to the residual-based one. This explains why the residual-based estimator of $\sigma_v^2$ works very well even when the sample size is small. Another advantage of the difference-based estimation is that it does not need to estimate the unknown function $g(\cdot)$. The asymptotic normalities of the difference-based estimators are established.


*MSC:* primary 62H12; secondary 62A10

*Keywords:* Repeated measurement; One-way error component; Residual-based method; Difference-based method; Asymptotic normality

## 1. Introduction

Consider a repeated measurement regression model

$$y_{ij} = g(x_i) + \varepsilon_{ij}, \quad i = 1, \ldots, n, \; j = 1, \ldots, m, \tag{1.1}$$

where $y_{ij}$'s are responses, $g(\cdot)$ is an unknown function, $x_i$'s are design points, $\varepsilon_{ij}$'s are random errors with a one-way error component structure, i.e.

$$\varepsilon_{ij} = \mu_i + v_{ij}, \tag{1.2}$$

where $\mu_i$ and $v_{ij}$'s are i.i.d random variables with mean zero, variance $\sigma_\mu^2$ and $\sigma_v^2$, respectively.

*Corresponding author. Tel.: +1 919 966 0229.
*E-mail address:* jyou@bios.unc.edu (J. You).

0167-7152/$ - see front matter 
doi:10.1016/j.spl.2006.11.020

Model (1.1)–(1.2) is a useful and flexible tool to fit multiple dependent variables. For the unknown function $g(\cdot)$, we can apply the usual kernel smoothing or series approximation procedure to estimate it consistently. Actually, except for $g(\cdot)$, the error variance $\sigma_\mu^2 = E(\mu_1^2)$ and $\sigma_v^2 = E(v_{ij}^2)$ are also important quantities. They describe the noise level and correlation between equations. Apart from the intrinsic interest as parameters of the model, estimators of them are essential in such tasks as the construction of efficient estimator of $g(\cdot)$, confidence regions, model-based tests, model selection procedures, signal-to-noise ratio determination, bandwidth selection and so on. Like Zhou and You (2005) we can construct estimators for them by estimated residuals. Suppose that $\hat{g}(\cdot)$ is a consistent estimator of $g(\cdot)$, for example the Gasser–Müller estimator. Based on $\hat{g}(\cdot)$ we can obtain the estimated residuals as

$$\hat{\varepsilon}_{ij} = y_{ij} - \hat{g}(x_i), \quad i = 1, \ldots, k, \; j = 1, \ldots, m.$$

Since $E(\varepsilon_{ij_1}\varepsilon_{ij_2}) = \sigma_\mu^2$ for $j_1 \neq j_2$ and $E(\varepsilon_{ij}^2) = \sigma_v^2 + \sigma_\mu^2$, from these estimated residuals we can construct estimators of $\sigma_\mu^2$ and $\sigma_v^2$, respectively, by

$$\hat{\sigma}_\mu^2 = \frac{1}{nm(m-1)} \sum_{i=1}^{n} \sum_{j_1=1}^{m} \sum_{j_2 \neq j_1} \hat{\varepsilon}_{ij_1} \hat{\varepsilon}_{ij_2} \quad \text{and} \quad \hat{\sigma}_v^2 = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \hat{\varepsilon}_{ij}^2 - \hat{\sigma}_\mu^2. \tag{1.3}$$

According to the simulation in Zhou and You (2005), although $\hat{\sigma}_v^2$ works very well $\hat{\sigma}_\mu^2$ works poorly, especially when the sample size is small.

In this paper, we propose a difference-based estimation for $\sigma_\mu^2$ and $\sigma_v^2$. We show the resulted difference-based estimator of $\sigma_\mu^2$ performs much better than the residual-based estimator $\hat{\sigma}_\mu^2$, especially when the sample size is small. In addition, we show the difference-based estimator of $\sigma_v^2$ is equal to the residual-based estimator $\hat{\sigma}_v^2$. This explains why the residual-based estimator of $\sigma_v^2$ works very well even in the case of small sample size. Compared with residual-based estimators, another advantage of the difference-based estimators is that they do not need to estimate the unknown function $g(\cdot)$ and avoid the selection of the bandwidth. The asymptotic normalities of the difference-based estimators are established. Difference-based estimation has been widely applied to estimate error variance. See, for example, Rice (1984), Gasser et al. (1986), Hall et al. (1990), Dette et al. (1998), Chen (2002) and so on. However, there are few results about the difference-based estimation used to replicate measurement models.

The layout of the remainder of this paper is as follows. In Section 2 we describe the proposed difference-based estimation. In Section 3 we present results from numerical studies. Section 4 gives the summary and all proofs of main results are relegated to the Appendix.

## 2. Difference-based estimation

Suppose that $\{y_{ij}, x_i, i = 1, \ldots, n, j = 1, \ldots, m\}$ is a sample from model (1.1)–(1.2). Similar to He et al. (2002), we assume that the sequence of designs $\{x_i\}_{i=1}^n$ is all scaled into the interval $[0, 1]$ and forms an asymptotically regular sequence (Sacks and Ylvisacker, 1970) in the sense that

$$\int_0^{x_i} p(x)\,\mathrm{d}x = \frac{i-1}{n-1}, \tag{2.1}$$

where $p(\cdot)$ denotes a positive density function on the interval $[0, 1]$.

Note that when $j_1 \neq j_2$

$$y_{ij_1} - y_{ij_2} = \varepsilon_{ij_1} - \varepsilon_{ij_2} = v_{ij_1} - v_{ij_2}$$

and

$$E(v_{ij_1} - v_{ij_2})^2 = 2\sigma_v^2.$$

Therefore, we propose a difference-based estimator of $\sigma_v^2$, which has the form

$$\check{\sigma}_v^2 = \frac{1}{2nm(m-1)} \sum_{i=1}^{n} \sum_{j_1=1}^{m} \sum_{j_2 \neq j_1} (y_{ij_1} - y_{ij_2})^2.$$