Combined effects of some chemicals on $\beta$–galactosidase activity using a new semiparametric errors in variables model

DOI: 10.7904/2068–4738–X(19)–42

Seçil YALAZ¹, Fatma MATPAN BEKLER²*, Ömer ACER²

¹Department of Statistics, Faculty of Science, Dicle University, 21280 Diyarbakır, TURKEY, ²Department of Biology, Faculty of Science, Dicle University, 21280 Diyarbakır, TURKEY, *Corresponding author: fmatpan@dicle.edu.tr

Abstract. The combined effects of some chemicals on $\beta$–galactosidase activity by *Anoxybacillus* sp. FMB1 were studied using semiparametric errors in variables methodology when the error has an unknown distribution. An experimental design was chosen to explain six chemicals, dithiothreitol (DTT), phenylmethylsulfonyl fluoride (PMSF), N–ethylmaleimide (NEM), iodoacetamide (Iod A), $\beta$–Mercaptoethanol ($\beta$–Mer) and 1,10–phenanthroline (phen) and to investigate the results. This technique constrained the number of real experiments performed while considering conceivable interactions between six chemicals. The optimal combinations of chemical concentration for maximum $\beta$–galactosidase activity was determined as 8 mM DTT, 8 mM PMSF, 8 mM NEM, 8 mM Iod A, 2 mM $\beta$–Mer and 2 mM phen. We also compared the no measurement error case and errors in variables case. The proposed estimator shows better performances.

Keyword: $\beta$–galactosidase, inhibition, semiparametric EIV Model, unknown distribution

Introduction

Beta–Galactosidase ($\beta$–D–galactoside galactohydrolase, lactase, E.C. 3.2.1.23) is a catalyst which catalyses glucose and galactose hydrolysis which are monosaccharides achieving lactose that is a disaccharide. This enzyme is of great importance to biotechnology with applications in dairy, food, pharmaceutical and medical industries [GEKAS and LOPEZ–LEYVA, 1985; LADERO et al., 2003; OAYYUM, 2010].

It has interactions with various chemicals in industrial studies where $\beta$–galactosidase is used and these affect the enzyme reaction.

Estimation of the effect of some chemicals by the traditional method requires changing one independent variable keeping the alternate factors steady.

The conventional methods for multifactor experimental design are time–consuming and incapable of detecting the true optimum, because of the interactions among the factors [LIU and ZENG, 1998].

Most studies used Response surface methodology (RSM) which is a collection of experiments, mathematical methods, and statistical inference that evaluates the combined effect of all the factors [ELBOL, 2003].

Tanyildizi and collab. and Abdel–Fattah and collab. reported the combined effects of macronutrients of media on enzyme production using response surface methodology [TANYILDIZI et al., 2005; ABDEL–FATTAH et al., 2013, BARBAT, et al., 2013].

In enzyme activity calculations, while linear effects of some chemicals on enzyme activity are observed, there is no such observation of some of them.

Jaspers and collab. used semiparametric mixture model to estimate nonparametric effects and linear effects on antimicrobial resistance [JASPERS et al., 2014].

The author is able to estimate the distribution taking all available information into account.

When the samples have errors stem from a deficiency of the measuring techniques the estimation may not always be an accurate reflection of reality.

In literature semiparametric errors in variables model has been mostly studied when the error distribution is known [LIANG, 2000].

However, in some applications, estimation of regression model when the independent variable is measured with an error drawn from an unknown distribution can be interested.
In this study, we used a new semiparametric error in variables model developed by Yalaz in 2015 in order to explain the combined effect of the chemicals on enzyme activity, considering that there is a measurement error in nonparametric variable and there is no information at all about the error distribution [YALAZ, 2015].

We also compared the optimal estimation of the proposed model (EIV) with the model when it is assumed that there is not a measurement error in variable of nonparametric part (NoME).

Taking into account measurement criteria and shape of nonparametric function, it has been seen that EIV estimator has better performances.

**Material and methods**

**Strain and medium**

Thermophilic *Anoxybacillus* sp. FMB1 (NCBI GenBank database accession number is KP992869) was obtained from Biology Department of Dicle University.

One mL of culture was inoculated in a 100 mL Erlenmeyer flask containing 25 mL Nutrient Broth (NB) medium composed of (g/L): beef extract, 10.0; peptone, 10.0; NaCl, 5.0 and incubated overnight at 55 °C for 24 h in a shaker.

**Enzyme analysis**

The crude extract was precipitated by ammonium sulphate added (on ice, with stirring) to obtain final ammonium sulphate concentration between 70 and 80 %. The precipitate was centrifuged and resuspended in 0.1 M sodium phosphate buffer (pH 9.0) and dialyzed against the same buffer.

After dialyze, the enzyme was concentrated by ultrafiltration cell (PBGC membrane, Millipore) and applied to a Sephadex G–75 column (1.5 x 30 cm) for gel permeation chromatography.

The protein was eluted with 0.1 M sodium phosphate buffer at a flow rate of 3 mL/min.

Fractions (8 mL) were collected and concentrated by ultrafiltration. The active pool from Sephadex G–75 chromatography was applied onto a column (PABTG–agarose) equilibrated with 50 mM sodium phosphate buffer (pH 6.0).

The fractions were measured at 280 nm for determination of amount of protein and enzyme assays were measured at 420 nm.

The active pools, which did not bind to the ligand on the affinity column, were dialysed in sodium phosphate buffer and concentrated by ultrafiltration and then assayed for enzyme activity and protein measurement.

β–galactosidase assay was determined according to Matpan Bekler and collab. 100 μL of enzyme were incubated at 60 °C for 10 min with 50 μL of 60 mM o–nitrophenyl–β–D–galactopyranoside (oNPG, Sigma, in 50 mM sodium phosphate buffer (pH 9.0)) [MATPAN BEKLER et al., 2015].

The enzyme reaction was terminated by the addition of 500 μL of 2 M sodium carbonate (Na₂CO₃) and the absorbance was measured at 420 nm.

Enzyme activity was expressed as o–nitrophenol (oNP) units liberated, where one unit (U) is defined as the amount of enzyme that released 1 μmol of oNP from oNPG per minute under the assay conditions (60 °C, pH 9.0).

An extinction coefficient (420 nm) for oNP of 4.3834 M⁻¹ cm⁻¹ was used to calculate the specific activity.

The enzyme content was determined by the Lowry method in 1951, method using bovine serum albumin (BSA) as a standard [LOWRY, 1951].

Specific activity is the amount of enzyme activity per milligram of protein (micromoles of product formed per minute per millgram of protein, or units per milligram).

**The effects of chemicals**

The effect of chelating agent (1,10–phenanthenol: phen) and chemicals (dithiothreitol: DTT), phenylmethylsulphonyl fluoride (PMSF), N–ethylmaleimide (NEM), iodoacetamide (Iod A) and β–Mercaptoethanol (β–Mer) on β–galactosidase activity was assayed using 1.5 mM oNPG solution in 0.1 M sodium phosphate buffer (pH 9.0) at 60 °C.
The phen was dissolved in methanol, and PMSF and NEM were dissolved in ethanol.

The partially purified enzyme was pre-incubated with all agents for 15 min at 60 °C and then enzyme activity was measured by spectrophotometer and enzyme unit was calculated.

Semiparametric errors in variables regression method

Regression analysis is based on some assumptions.

The most important assumption is that the shape of the relationship between dependent and independent variables is known.

Because estimates which are made in cases where the assumptions would not be provided will not be good estimates, nonparametric and semiparametric regression methods which can stretch the assumption of linearity assumption is needed.

In cases where the number of independent variables is more than two, semiparametric regression method is preferred due to the difficulty in making the analysis and interpreting the graphics done by nonparametric methods.

Semiparametric regression model defined in equation (1) summarizes complex data sets in an understandable form, ignoring the trivial details and retains the important features of the data and thus implementing substantive decisions:

\[ y = X^T \beta + g(x^*) + \Delta y \] (1)

where \( y \) is dependent variable, \( x = (x_1, ..., x_p) \) is a random vector of independent variables, \( \beta \) is a vector of unknown parameters and \( x^* \) is a random variable defined in \([0,1]\), \( g \) is an unknown function and \( \Delta y \) is model error has zero mean and \( \alpha^2 \) variance (see Robinson) [ROBINSON 1988].

Regression models with measurement error in the independent variables are known as errors in variables that are known models or measurement error models described equation (2) should be used instead of the models based on standard assumptions that lead to inconsistent estimates when used in this circumstance:

\[ \chi = x^* + \Delta x \] (2)

where \( \chi \) is variable measured with error, and \( \Delta x \) is the measurement error [JIANG 2000].

Yalaz presents a detailed answer to the question that how the estimations of regression functions and densities can be obtained if the measurement error has an unknown distribution in a semiparametric regression model [YALAZ, 2015].

They accomplish the identification of the density of an unobserved random variable when the joint density of two error-contaminated measurements of that variable is known.

They also derived asymptotic normality of proposed estimator. With the simulation study, they show that the resulting rates are comparable to Nadaraya–Watson estimators using kernel deconvolution approach, which provide consistent estimation under the much stronger assumption that the density of the measurement error is known, and no measurement error case.

The model is described as

\[ y = X^T \beta + g(x^*) + \Delta y, \]

\[ \chi = x^* + \Delta x \]

(3)

Where \( z \) is variable which describes second imperfect measurement of \( x^* \), and \( \Delta z \) is the measurement error.

Under the assumptions described by Yalaz generalized that the squares estimator of \( \beta \) can be indicated in terms of moments that involve the observable variables

\[ x, y, z \] [YALAZ, 2015] as

\[ \hat{\beta} = (X^T X)^{-1} (X^T y) \] (4)

where \( y = (y_1, ..., y_n) \) with \( y_i = Y_i - \omega_x(x^i, h) \)

and \( \bar{x} = (x_i^*, ..., x_n^*) \) with \( \bar{x}_i = x_i^* - \omega_x(x^*, h) \)

where

\[ \omega_x(x^*, h) = \frac{\bar{x}_i(x^*, h)}{m_x(x^*, h)} \]

and where for \( t = x_i^* \)

\[ \omega_x(x^*, h) = \frac{1}{2 \pi} \int -i \xi x^* \phi_x(\xi) \exp(-i \xi t) d\xi \] (6)

where \( \phi_x(\xi) \equiv E[\exp(i \xi x^*)] \) is given by

\[ \phi_x(\xi) = \phi_0(\xi) m_x(\xi) / m_1(\xi) \] (7)
where $l = \sqrt{-1}$ and $k(\xi)$ is the Fourier transform of the kernel $k(x^*)$, $h$ is the bandwidth parameter.

Theoretically, if $\beta$ were known absorbing into $y$; the estimate of $\beta(x^*)$ at a given point $x^*$ is

$$\hat{\beta}(x^*, h) = \frac{M_4(x^*, h)}{M_2(x^*, h)}$$

where

$$M_4(x^*, h) = \frac{1}{2\pi} \int k(\xi) \varphi_4(\xi) \exp(-i\xi x^*) d\xi$$

and

$$M_2(x^*, h) = \int k(\xi) \varphi_2(\xi) \exp(-i\xi x^*) d\xi$$

The bandwidth can be selected using cross validation.

In this paper, firstly we decided linear and nonparametric relationship with dependent and independent variables, secondly we used the method described by Yalaz using the relations between independent variables [YALAZ, 2015].

As shown in Figure 1 while $x_1$:DTT, $x_2$:PMSF, $x_3$:Iod A, $x_4$:β–mer, and $x_5$:Phen, $x^*$:NEM, have a linear relation with dependent variable $y$ : β–galactosidase activity, $x^*$:NEM has nonparametric relation with $y$. $x_4$:β–mer is admitted to have a linear relationship with β–galactosidase activity based on the reason that after a certain delay the relationship behaves linear.

Figure 1 also shows that one of independent variable called NEM has a strange structure.

Results and discussion
The main goal of this study was to reveal the effect of concentrations of some chemicals on β–galactosidase activity and for that purpose experimental design in semiparametric regression method was used as variable of nonparametric part that had unknown measurement error.

To decide the linear and nonparametric relationship with dependent and independent variables we used in the Figure 1.

Since we are concerned with the analysis of data produced from the experiments, we use experimental design to take time and effort to organize the experiment properly to ensure that the right type of data, and enough of it, is available to answer the questions of interest as clearly and efficiently as possible.
Hence, nonparametric variable could be seen as it may have measurement error.

In order to explain the combined effect of the chemicals on enzyme activity, considering that there is a measurement error in nonparametric variable and there is no information at all about the error distribution we used the method described in the previous section. Because we don’t have any information about the distribution of error, we used two imperfect measurements of $x^*$ which are obtained by two different experiments and called as $x_1$ and $x_2$.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Symbol coded</th>
<th>Ranges and levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
<td>$x_1$</td>
<td>$x_2$</td>
</tr>
<tr>
<td>DTT</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>PMSF</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Iod A</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>β–mer</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Phen</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>NEM</td>
<td>$x^*$</td>
<td>2</td>
</tr>
</tbody>
</table>

The ranges and levels of six independent variables are presented in Table 1.

The activities of the chemicals within the range of 1–10 mM on partially purified enzyme have been calculated.

Experiments were planned to obtain 26 trials, plus 2x6 star configuration and a centre point.

The ranges and levels of six independent variables are presented in Table 1.

We write codes in MATLAB to produce the regression equation above and performance measuring criterion below [MATLAB 2015].

To compare the no measurement error case (NoME) and errors in variables case (EIV), we used the coefficient determination ($R^2$)


\[ R^2_{EIV} = 0.9102, R^2_{NoME} = 0.9095 \]

This performance measuring criterion shows that proposed model has better performance.

### Table 2.

<table>
<thead>
<tr>
<th>Run order</th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th>X5</th>
<th>X*</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1447.48</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1446.627</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1308.489</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1306.549</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1448.395</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1447.549</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1439.896</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1307.971</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1470.476</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1469.781</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1337.488</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1335.794</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1471.316</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1470.628</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1338.785</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1337.104</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1453.583</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1452.769</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1315.681</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1313.797</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1454.48</td>
</tr>
<tr>
<td>22</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1453.674</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1317.066</td>
</tr>
<tr>
<td>24</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1315.196</td>
</tr>
<tr>
<td>25</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1476.4</td>
</tr>
<tr>
<td>26</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1475.743</td>
</tr>
<tr>
<td>27</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1344.442</td>
</tr>
<tr>
<td>28</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1342.8</td>
</tr>
<tr>
<td>29</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1477.224</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1476.574</td>
</tr>
<tr>
<td>31</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1345.717</td>
</tr>
<tr>
<td>32</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1344.088</td>
</tr>
<tr>
<td>33</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1471.884</td>
</tr>
<tr>
<td>34</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1471.207</td>
</tr>
<tr>
<td>35</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1340.89</td>
</tr>
<tr>
<td>36</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1339.243</td>
</tr>
<tr>
<td>37</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1472.711</td>
</tr>
<tr>
<td>38</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1472.04</td>
</tr>
<tr>
<td>39</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1342.16</td>
</tr>
</tbody>
</table>

In Figure 2 we also demonstrate the nonparametric functions of compared models in terms of the orders (Figure 2a) and real data in terms of \( x^* \) (Figure 2b).

This shows that our model (EIV) is closer to real data.

The significant achievement of the present work was to enable us to have no information about a measurement error in nonparametric variable and the combined effect of the six chemicals on the enzyme activity.

This study is significant in that the results showed us that without an experiment in laboratory conditions it is possible to estimate the optimum chemical combination and concentration for optimum enzyme activity using described model.

The optimal combinations of chemicals concentration for maximum \( \beta \)-galactosidase activity was determined as 8 mM DTT, 8 mM PMSF, 8 mM NEM, 8 mM Iod A, 2 mM \( \beta \)-Mer and 2 mM phen.
It is known that Cys residue is modified by PMSF and NEM. In our experiment, the enzyme activity increased in the presence of PMSF but the results were not linear for NEM.

Therefore, we thought that it could stem from a measurement error and we tried to explain the effect of NEM in total results using semiparametric errors in variables model.

The enzyme activity was slightly activated by NEM.

In addition, thiol compounds such as β-mercaptoethanol and DTT caused stimulatory effect on enzyme activity indicating that cysteine residue(s) do not take part in catalysis.

In addition, this activation is attributed to the reduction in aggregate size by destroying the intermolecular disulfide linkages and protection of thiol groups that stabilize the three-dimensional structure of enzyme [Khedler et al. 2008].

![Figure 2a. The nonparametric functions of compared models (NoME with dashed line and EIV with solid line).](image)

![Figure 2b. Graphs of the nonparametric functions of compared models and real data in terms of x^*](image)

The enzyme activity was slightly activated by divalent metal chelator 1,10-phenanthroline a combined effect while the enzyme activity was significantly inhibited by phen a single effect of in the total activity result.
Phen targets mainly zinc metallopeptidases, with a much lower affinity for calcium [SALVESEN and NAGASE, 2001]. This result indicates that divalent metal ion cofactors are necessary for the activity of the β-galactosidase or in other words the enzyme is a metalloenzyme.

Similarly, the enzyme activity was enhanced by DTT. Patel and collab. have reported that lactase activity was greatly stimulated by reducing agents such as DTT [PATEL et al., 1985].

It is known that iodoacetamide is the reagent alkylating SH group.

Conclusions

In our results, either single or combined effect of IodA on the enzyme activity has decreasing influence. The method that was selected for the optimization of chemical concentration was found to be quite simple, adequate, and taking less time and material.

This study showing the estimation of the combined effect of various inhibitors and chemicals on enzyme activity further clarifies the nature of the active site of β–galactosidase, which is a member of the GH42 family.

References

2. Barbat, C.; Rodino, S.; Petrache, P.; Butu, M.; Butnariu, M. Microencapsulation of the allelochemical compounds and study of their release from different, Digest journal of nanomaterials and biostructures, 2013, 8(3), 945–953.


Received: December 07, 2018
Article in Press: April 30, 2019
Accepted: Last modified on: May 20, 2019