Extrapolation estimation for parametric and nonparametric regression with normal measurement error

by

Kanwal Ayub

B.S., Lahore University of Management Sciences, 2012

M.S., Western Illinois University, 2016

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Abstract

For the general parametric and nonparametric regression models with covariates contaminated with normal measurement errors, this dissertation proposes an accelerated version of the classical simulation extrapolation algorithm to estimate the unknown parameters in the parametric as well as the nonparametric regression functions. For the parametric regression model, the proposed algorithm successfully removes the simulation step of the classical SIMEX algorithm by applying the conditional expectation directly to the target function thereby generating an estimation equation either for immediate use or for extrapolating, thus significantly reducing the computational time.

For the nonparametric regression models with covariates contaminated with normal measurement errors, the regression functions are estimated by applying the conditional expectation directly to the kernel-weighted least squares of the deviations between the local linear approximation and the observed responses, thereby successfully bypassing the simulation step needed in the classical simulation extrapolation method, hence significantly increasing the computational efficacy. It is noted that the proposed method also provides an exact form of the extrapolation function, but the extrapolation estimate generally cannot be obtained by simply setting the extrapolation variable to negative one in the fitted extrapolation function if the bandwidth is less than the standard deviation of the measurement error.

Large sample properties of the proposed estimation procedures, including the consistency and the asymptotic normality, are thoroughly discussed. Potential applications of the proposed estimation procedures are illustrated by examples, simulation studies, as well as a real data analysis.

Extrapolation estimation for parametric and nonparametric regression with normal measurement error

by

Kanwal Ayub

B.S., Lahore University of Management Sciences, 2012

M.S., Western Illinois University, 2016

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Approved by:

Major Professor
Weixing Song

Copyright

Abstract

For the general parametric and nonparametric regression models with covariates contaminated with normal measurement errors, this dissertation proposes an accelerated version of the classical simulation extrapolation algorithm to estimate the unknown parameters in the parametric as well as the nonparametric regression functions. For the parametric regression model, the proposed algorithm successfully removes the simulation step of the classical SIMEX algorithm by applying the conditional expectation directly to the target function thereby generating an estimation equation either for immediate use or for extrapolating, thus significantly reducing the computational time.

For the nonparametric regression models with covariates contaminated with normal measurement errors, the regression functions are estimated by applying the conditional expectation directly to the kernel-weighted least squares of the deviations between the local linear approximation and the observed responses, thereby successfully bypassing the simulation step needed in the classical simulation extrapolation method, hence significantly increasing the computational efficacy. It is noted that the proposed method also provides an exact form of the extrapolation function, but the extrapolation estimate generally cannot be obtained by simply setting the extrapolation variable to negative one in the fitted extrapolation function if the bandwidth is less than the standard deviation of the measurement error.

Large sample properties of the proposed estimation procedures, including the consistency and the asymptotic normality, are thoroughly discussed. Potential applications of the proposed estimation procedures are illustrated by examples, simulation studies, as well as a real data analysis.

Table of Contents

List of Figures

List of Tables

Acknowledgments

Firstly, I would like to thank the One who made me, guided me and gave me everything I have in this world.

Second, I want to thank my parents and my aunt, my sister and my brother for their utmost love and support. My mother especially for everything she does for me not to mention all the times she stays up at night praying for me and also listening to my complaints. My father for checking up on me to make sure I'm okay and for being my constant support. My sister for bearing with me for the past several years and always being there for me and forgiving me despite all my short-comings. Words are not enough to describe everything my family has done for me and how grateful I am to them and I am indebted to them since I know I can never pay them back for the time and effort they have invested in me.

I also majorly want to thank my advisor, Dr. Weixing Song, to whom this degree is dedicated to. If it wasn't for Dr. Song, I would have left the program before I even started it. I just want to thank him for agreeing to take me as his student and for being my constant guide and mentor throughout my time here. He has taught me the most in these past 4 years than anybody else could in a lifetime. I have learnt not just statistics from him but also kindness, humility and just to be a good person in general. Thank you Dr. Song for your patience, your sincerity, your kindness and your humbleness. I wish there were more people like you or atleast have the same amount of dedication that you have towards your profession. Also, there isn't enough space in here for me to express how honored I am to have been your student and have learnt from you or how grateful I am for everything you have done for me. Also, I would like to apologize if I ever let you down or didn't come up to your expectations.

I would also like to thank Dr. Cen Wu, Dr. Gyuhyeong Goh and Dr. Zhoumeng Lin for agreeing to be my committee members. And finally I would like to thank the Department of Statistics for having accepted me as a student and the Graduate School at Kansas State University for supporting me when things got tough.

Chapter 1

Introduction

To make efficient statistical inferences in a regression analysis, it is imperative to estimate the parameters in the regression function precisely. However, in real applications, it is often the case that the explanatory variables cannot be observed directly. Instead, the covariates observed are contaminated with measurement error. One might assume that measurement error is the same as the 'noise' seen in the observed values ('Y') in a regression model. However, measurement error is witnessed in the explanatory variables ('X'), and therefore, the observed quantities are not an accurate reflection of the true regressors. The source of measurement error could be due to random errors (noise) or systematical errors (statistical bias).

Errors-in-variables settings are common in disciplines such as biology, epidemiology, and clinical assessments, to name a few. Various reasons can cause measurement error – the most common being the inaccuracy introduced by instrumental apparatuses. Other causes include bias from self-reported measures, inherent biological variability, and the exorbitant costs of extracting exact information.

A decent example of the occurrence of measurement error is blood pressure readings of individuals. Blood pressure has daily as well as seasonal variation, so the blood pressure reading of an individual taken at one time would not only be different six months later but could also be different six hours later. Therefore, any dataset that uses blood pressure as an independent/explanatory variable would need to account for these variations. Hence, models fit to such datasets fall under the category of measurement error models.

Fitness trackers are a great example of where measurement errors are rife. Steps measured using trackers of different brands usually yield different results. Furthermore, trackers from the same brand would also result in mismatched observations demonstrating the variability in the number of steps taken by an individual due to measurement error.

Another example is that of the measurement of soil nitrogen levels. Now soil nitrogen varies not just by location but also by depth. So, for the same plot of land, not only would the soil nitrogen fluctuate at different sites within the same field but would also be distinct at various depths for the exact location. Furthermore, separate labs would yield distinct levels of nitrogen for identical samples of soil and lab technicians would add to the source of variability. Again, any model that uses nitrogen levels in soil as an explanatory variable would then need to account for the errors in variables.

Dietary assessment using food frequency questionnaires is another example where measurement error is rampant. For instance, using an FFQ to measure daily caloric intake results in the logged calories not being an exact measure of the actual calories consumed. This could be because, in general most people do not keep track of what they consume daily, and diets generally change on a monthly if not weekly basis. Also, some people are embarrassed about the amount of snacking they do, so they tend to underreport the calories they consume. This leads to FFQs being laced with measurement error, which results in quantities such as Vitamin D intake or protein consumption not being measured precisely. Hereafter, any models based on dietary assessments using FFQs must adjust for errors-in-variables.

In standard regression models, the only errors accounted for are those in the dependent variables, a.k.a responses. Regressors are understood to have been quantified precisely; that is, they are assumed to be the true reflection of the variable of interest. Therefore, estimations based on the conventional assumptions beget inconsistent parameter estimates. In the case of simple linear regression, classical measurement error results in the coefficient being under-estimated (this is also known as 'attenuation bias').

The reason why measurement error is dangerous is that not only does it cause biased parameter estimation but also leads to loss of power, thereby mitigating the chances of detecting clinically meaningful relationships between variables. In the case of nonlinear models, features of the data are masked; hence graphical representations of the data are not a true reflection of how the response variable is distributed with respect to the explanatory variable. This was referred to as the 'triple whammy' by Carroll et al. (2006). Besides causing bias in parameter estimates and loss of power, measurement errors also lead to inaccurate coverage of confidence intervals.

The classical error framework is the simplest of the errors-in-variables models. In the measurement error context, we denote X to be the unobserved, true explanatory variable and Z to be the observed variable. In the classical error model Z is related to X by $Z = X + U$, where U is a mean zero random variable independent of X and variance equal to σ_u^2. The classical measurement error model is additive and unbiased with $E[Z|X] = X$.

Another type of error-in-variable is the Berkson error, which relates X with Z by $X = Z + U$, where U is again a random variable independent of Z with a mean of zero and variance equal to σ_u^2. In the Berkson error model, Z is unbiased for X since $E[X|Z] = Z$.

Some frequently used procedures in measurement error modeling include Regression Calibration (denoted RC for short) and Deconvolution density estimation. Because of the ease of its applicability in current software, RC is most frequently used when adjusting for measurement error in covariates. The idea behind RC is the estimation of X (the unobserved, true variable) on Z (the observed, error-contaminated value). The resulting estimates are used to replace the unknown values of X in the model, and standard analysis is then performed. When using RC, some assumptions need to be met to achieve consistent or approximately consistent estimators of the parameter of interest - two of them being that X and Z follow a linear relationship that is also homoscedastic. If these conditions are not satisfied, then the resulting estimate could be misleading.

To ensure the identifiability of parameters, some additional information is needed. This could be, for example, more data that is beyond the original study sample. For instance, a subgroup of subjects from the original study can be used to record observations from X. This is called the internal validation data set. Another way would be to collect replicate observations of X, which would be in the same way observations on Z were recorded.

The deconvolution procedure entails estimating the density function of the unobserved variable X. This can be done if the density function of U (the measurement error) is known and its characteristic function is valid on the entire real line. Then the characteristic function of X can be estimated by the ratio of the characteristic functions of Z and U, where the characteristic function of Z is determined using the observations on Z. The deconvolution kernel density estimate is then defined to be the inverse Fourier transform of the ratio of the estimated characteristic function of Z with that of U.

The mathematical intricacy and complex technical details make the deconvolution procedure difficult to apprehend for applied researchers and cause it to appear redundant when dealing with relatively uncomplicated models. In contrast, the SIMEX approach as proposed by Cook and Stefanski (1994) provides a much easier simulation-based approach to lessen the bias in parameter estimates of measurement error models.

In the original work of Cook and Stefanski (1994), the SIMEX method was introduced to correct for the attenuation caused by measurement errors. The simulation method is useful in cases when the measurement error variance is known or can be estimated in a reasonable manner. It involves simulating errors in increments, adding them to the error-contaminated data, and estimating parameters based on the new pseudo-data. A trend of parameter estimates versus the increments in measurement error is established, and the trend is extrapolated to the case of no measurement error.

Besides the ease of implementation of this method is the added advantage that a graphical representation can be done whereby the effect of measurement error on the parameter estimate can be seen. This is of great benefit for people who are not well-versed in the field of the statistical theory of measurement error models but are otherwise knowledgeable of standard statistical methods. Also, the fact that this method is quite general allows for its wide applicability as well as in situations where a novel model is under consideration and conventional methods have not been methodically examined.

The method was initially developed to fit error-contaminated data by using nonstandard generalized linear models. The methods that existed prior to this procedure were not only complex and required specialized software but also entailed approximations, a fact that made them inconvenient to authenticate intricate models.

Suppose $Z_i = X_i + U_i$ where X_i is a vector of model covariates and U_i is the measurement error. Let Y_i be the response variable with Y_i depending on X_i through some unknown parameter θ. Let $(Z_i, Y_i), i = 1, ..., n$, be the observed data and assume U_i to be multivariate Gaussian with zero mean and covariance matrix Σ_u. Denote T to be the estimator of θ from the observed data. Then $\hat{\theta}_{true} = T(X_i, Y_i)_{i=1,\dots,n}$ would be the true estimator had the true covariates X_i been observed. Since X_i are not observed, we use the data with measurement error to calculate the naive estimator, $\hat{\theta}_{naive} = T(Z_i, Y_i)_{i=1,\dots,n}$, treating Z_i as if it contains no measurement error. Naive estimators are usually not consistent nor unbiased for θ. SIMEX is a popular method used to deal with any biases that result in models with errors-in-variables. The intensity of the effect of measurement error on the parameter of interest is evaluated through simulation, and the estimator is determined through extrapolation of the simulation results. SIMEX is implemented as follows:

1. Simulation Step
   (a) A grid of λ values is chosen such that $\lambda = 0 < \lambda_1 < ... < \lambda_M$
   (b) For each λ_m:
      i. B sets of pseudodata are generated by adding a random error to $Z_i, i = 1, ..., n$, that is,

$$Z_i^{(b)}(\lambda_m) = Z_i + \sqrt{\lambda_m} U_i^{(b)} \tag{1.0.1}$$

      where $b = 1, ..., B$ with $U^{(b)} \sim N_p(0, \Sigma_u)$.
   (c) The naive estimator is calculated for each set of pseudodata, that is, $\theta^{(b)}(\lambda_m) = T(Z_i^{(b)}(\lambda_m), Y_i)_{i=1,\dots,n}$
   (d) An average of the estimators is taken, that is, $\hat{\theta}(\lambda_m) = \frac{1}{B}\sum_{b=1}^{B} \theta^{(b)}(\lambda_m)$
2. Extrapolation Step:

(a) $\hat{\theta}(\lambda)$ is modeled as a function of λ.

(b) The model is extrapolated back to $\lambda = -1$ to obtain $\hat{\theta}_{simex}$.

Heuristically, $Z_i^{(b)}(\lambda_m)$ is obtained by increasing the measurement error in Z_i by adding a new random error $\sqrt{\lambda_m} U_i^{(b)}$ to Z_i. The pseudodata $(Z_i^{(b)}(\lambda_m), Y_i), i = 1, ..., n$, is then used to calculate the naive estimator, that is, $\theta^{(b)}(\lambda_m) = T(Z_i^{(b)}(\lambda_m), Y_i)_{i=1,...,n}$. The variability due to the simulated errors for a fixed value of λ_m is reduced by increasing the number of sets of pseudodata generated and then averaging over the naive estimators from all the sets of pseudodata to obtain $\hat{\theta}(\lambda_m) = \frac{1}{B}\sum_{b=1}^{B} \theta^{(b)}(\lambda_m)$. The covariance matrix of the measurement error component for each set of pseudodata is given by $(1+\lambda)\sigma_u$ and so extrapolating to $\lambda = -1$ results in the case where there is no measurement error. This estimator is therefore called $\hat{\theta}_{simex}$.

Popular choices of extrapolation functions are:

- The linear function $\Gamma_{lin}(\lambda) = \gamma_0 + \gamma_1 \lambda$
- The quadratic function $\Gamma(\lambda) = \gamma_0 + \gamma_1 \lambda + \gamma_2 \lambda^2$
- The nonlinear function $\Gamma_{nonlin}(\lambda) = \gamma_0 + \frac{\gamma_1}{\gamma_2 + \lambda}$

It should be noted that the simulation component of SIMEX is computationally quite expensive, which renders the method inadequate, especially in situations where large datasets are used to establish relationships between variables. Furthermore, SIMEX can only be used when the measurement error is from a normal distribution (as noted by Koul and Song (2014)), with its variance being either known or reasonably estimated. It is imperative to note that the extrapolation step needs to be performed in a convincing manner. Cook and Stefanski (1994) pointed out in their paper that the linear, quadratic and, nonlinear extrapolation functions are exact for certain estimators when the distribution of the measurement error is normal with the asymptotic bias being of order $O(\sigma^6)$ when the quadratic and nonlinear extrapolation functions are used and $O(\sigma^4)$ when the extrapolation function is linear. Koul and Song (2014) proved that the SIMEX procedure only works for the case of Normal measurement error and showed that when the parameter of interest involves the fourth moment of the data, then the SIMEX estimator is not robust when the measurement error non-normally distributed. They introduced a method called the L-SIMEX, which was specifically designed for the case when the measurement error follows a Laplace distribution and provided theoretical and empirical evidence to justify the procedure. The L-SIMEX method follows the original SIMEX methodology in the sense that it uses the 'addition' component of the OG SIMEX approach in the simulation step and also follows the extrapolation step. It is noted in the paper that unless the exact relationship between λ and $\hat{\theta}(\lambda)$ is known, a scatter plot of $(\lambda_j, \hat{\theta}(\lambda_j)), j = 1, 2, \ldots, m$ can be used to determine the trend of $\hat{\theta}(\lambda)$ with respect to λ and a least squares procedure can be used to determine the case when $\lambda = -1$. Again, it is noted that like the N-SIMEX, the L-SIMEX suffers from the same dilemma in that extrapolation provides only an approximation since the true extrapolation function is not really known for most cases. The accuracy of the approximation relies heavily on the Monte Carlo error in the simulation step, but this can be improved by increasing the sample size and the number of pseudo-datasets generated. However, this results in increased computational time. Like the N-SIMEX, the linear, quadratic, and nonlinear extrapolation functions are also used in the L-SIMEX, with the quadratic and nonlinear extrapolants having an asymptotic bias of the order $O(\sigma_u^6)$.

Carroll et al. (1999) considered the problem of estimating parameters in a non-parametric regression function in the presence of errors in predictor variables. In the non-parametric setup, the regression of a response Y on a covariate X is given by $E(Y|X) = m(x)$. Again, X is not observed but what is observed is an error-contaminated value called Z, related to X by an additive error model, $Z = X + U$ with U being from a normal distribution with a mean of zero and a constant variance (assumed to be known). The problem here again lies in estimating $m(.)$, since observations on X are not known and the only available data is that of Y and Z. Fan and Truong (1993) showed that when $m(.)$ has k derivatives with the measurement error coming from a normal distribution with a known variance, then for a sample of size n, the fastest rate of convergence of a nonparametric estimator of $m(.)$ is $\log(n^{-k})$. However, most if not all practical progress in the field of measurement error for nonlinear models has been through the implementation of approximately consistent estimators, that is, estimators that converge in probability to some constant that is approximately equal to the parameter being estimated. When the measurement error variance is small, naïve estimators have a bias of order $O(\sigma_u^2)$, whereas the approximately consistent estimators have a bias of $O(\sigma_u^6)$ or less. In the nonparametric regression setup, B is chosen to be a large integer that is also finite (usually chosen to be between 50 and 200). Estimation of $E(Y|X)$ at x_0 is then considered. For a $\lambda_k > 0$ and $b = 1, \ldots, B$, a set of independent, normal random variables $V_{i,b}$ is generated from a Normal distribution with zero mean and variance σ_u^2 and a new set of pseudo-data is generated by adding the errors to Z_i and the result denoted by $Z_{i,b}(\lambda_k)$. Therefore, $Z_{i,b}(\lambda_k) = Z_i + \sqrt{\lambda_k} V_{i,b}$. This pseudo-data along with the Y_i's are used to estimate $\hat{m}_{b,\lambda}(x_0)$. The average of all the estimates over $b = 1, \ldots, B$ is denoted by $\hat{m}_\lambda(x_0)$. The non-parametric SIMEX estimator is then determined via a three-step process: In the first step, a grid of $\lambda' s$ is chosen so that $\lambda = 0 < \lambda_1 < \lambda_2 < \ldots < \lambda_k$ and $\hat{m}_\lambda(x_0)$ is computed for each λ; for the second step, a function is fit to model $\hat{m}_\lambda(x_0)$ versus λ (one can use either a linear, quadratic, or non-linear function); in the third step, the model is extrapolated to $\lambda = -1$ to obtain the SIMEX estimate of $m(x_0)$.

As is obvious, the SIMEX procedure in the non-parametric setup is quite like in the parametric setup, with the biggest difference being that the local linear estimator is used to evaluate $m(.)$, since an exact functional is not assumed. Even though the bias of the approximately consistent estimator using the SIMEX method is small compared to methods that ignore measurement error in the nonlinear setup, the computational intensity of the SIMEX procedure renders the method inadequate in practice, specifically when large datasets are at stake.

This dissertation consists of two major topics. In the second chapter, for the general parametric regression models with covariates contaminated with normal measurement errors, an alternative method to the traditional simulation extrapolation algorithm is proposed to estimate the unknown parameters in the regression function. By applying the conditional expectation directly to the target function, the proposed algorithm successfully removes the simulation step, by generating an estimation equation either for immediate use or for extrapolating, thus providing a possibility of reducing the computational time or the Monte Carlo simulation error. Large sample properties of the resulting estimator, including the consistency and the asymptotic normality, are thoroughly discussed. Potential wide applications of the proposed estimation procedure are illustrated by examples, simulation studies, as well as a real data analysis. A manuscript based on this chapter is published as the journal paper Ayub et al. (2022).

In the third chapter, for the nonparametric regression models with covariates contaminated with normal measurement errors, we proposes an extrapolation algorithm to estimate the regression functions. By applying the conditional expectation directly to the kernel-weighted least squares of the deviations between the local linear approximation and the observed responses, the proposed algorithm successfully bypasses the simulation step in the classical simulation extrapolation, thus significantly reducing the computational time. It is noted that the proposed method also provides an exact form of the extrapolation function, but the extrapolation estimate generally cannot be obtained by simply setting the extrapolation variable to negative one in the fitted extrapolation function, if the bandwidth is less than the standard deviation of the measurement error. Large sample properties of the proposed estimation procedure are discussed, as well as simulation studies and a real data example being conducted to illustrate its applications.

Chapter 2

Extrapolation Estimation for Parametric Regression with Normal Measurement Error

2.1 Introduction

As a simulation-based estimation method, the classical simulation-extrapolation (SIMEX) algorithm has enjoyed much popularity among its peers in the measurement error literature. Suppose a random variable X is generated from a population whose distribution is characterized by an unknown parameter θ. Unfortunately, X cannot be directly measured, instead, what we can observe is a surrogate variable Z. A commonly used structure in measurement literature assumes that Z and X are related through $Z = X + U$, where U is called the measurement error, and is independent of X. We further assume that U has a normal distribution with mean 0 and variance σ_u^2 which is assumed to be known. If X can be observed, suppose a statistic $T(\boldsymbol{X})$ can be found to estimate θ based on a sample $\boldsymbol{X} = (X_1, \ldots, X_n)$ of size n from X.

Now suppose we have the data $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$. A typical classical SIMEX procedure for estimating θ consists of the following three steps. In the first step, n i.i.d. random numbers V_i's are generated from $N(0, \sigma_u^2)$, and for a pre-specified nonnegative number λ, $\tilde{Z}_i(\lambda) = Z_i + \sqrt{\lambda} V_i$, $i = 1, 2, \ldots, n$ are calculated, and based on these pseudo-data, $T(\tilde{\boldsymbol{Z}}(\lambda))$ is calculated, where $\tilde{\boldsymbol{Z}}(\lambda) = \{\tilde{Z}_1(\lambda), \ldots, \tilde{Z}_n(\lambda)\}$. This operation is repeated for a large number of times, say B. Denote the resulting T-values as $T_b(\tilde{\boldsymbol{Z}}(\lambda))$, $b = 1, 2, \ldots, B$. Finally, averaging these B quantities concludes the first step, and we denote the resulting average by $\overline{T}(\lambda)$; In the second step, repeating the first step for a sequence of nonnegative λ values, for example, $\lambda = \lambda_1, \ldots, \lambda_K$, we obtain $\overline{T}(\lambda_1), \ldots, \overline{T}(\lambda_K)$. In real applications, a rule-of-thumb for choosing the sequence of values $\lambda_1, \ldots, \lambda_K$ is to select K equally spaced values from $[0, 2]$ for K around 20. In the last step, a trend of $\overline{T}(\lambda)$ with respect to λ is identified, the trend is then used to evaluate the value of $\overline{T}(\lambda)$ at $\lambda = -1$, and the extrapolated value $\overline{T}(-1)$ is the SIMEX estimate of θ. For more information on the SIMEX procedure, see the seminal papers of Cook and Stefanski (1994) and Stefanski and Cook (1995). The asymptotic properties of the SIMEX procedure, when σ^2 is small, is investigated in Carroll et al. (1996). Although we briefly introduce the classical SIMEX procedure for the univariate X cases, the multivariate scenarios are accommodated very well.

In all the literature involving applications of the SIMEX procedure, the three steps described above are strictly followed. In particular, pseudo data are always generated to provide the values of $T(\tilde{\boldsymbol{Z}}(\lambda))$, and an average follows. It is well known that the simulation step in the SIMEX procedure is notoriously time-consuming in the SIMEX procedure, in particular, when the estimator $T(\boldsymbol{X})$ has no closed-form and is determined inexplicit by an optimization process. The other drawback in using the classical SIMEX approach is the choice of the extrapolation function in the extrapolation step. Except for some very special cases, there are no tractable extrapolation functions to use in general. In fact, in most applications, the approximate ones, such as the linear, quadratic, and nonlinear forms, are taken as the working extrapolation functions. In this paper, we will construct an estimation procedure to avoid, or at least partially avoid, these drawbacks in the classical SIMEX procedure. Finally, it is also worth to mention that many papers indicate the classical SIMEX procedure is robust to the distributional assumption on the measurement error, that is, even when the measurement error is not normally distributed, the SIMEX procedure still provides reasonable estimates. However, Koul and Song (2014) theoretically proved that this is not true in general. Nevertheless, throughout this paper, we shall assume U has a normal distribution.

This paper is organized as follows. Some examples that motivate our research, such as the linear, quantile, and expectile regressions etc., are introduced in Section 2.2; the proposed extrapolation estimation procedure for the general parametric regression model is constructed in Section 2.3, together with an exploration on the large sample results. A discussion on the extrapolation function can also be found in this section; numerical studies are conducted in Section 2.4; Section 2.5 includes a discussion on the potential extension of the proposed estimation procedure to some semi-parametric regression models, as well as the robustness of the proposed estimation procedure to the normality assumption. All the technical proofs of the theoretical results are deferred to Appendix A.

2.2 Motivating Examples

Before formulating our extrapolation estimation procedure for the general parametric regression models, we start with a very simple example to see how our research idea has been developed. Suppose we have a simple linear errors-in-variables regression model $E(Y|X) = \alpha + \beta X$, and $Z = X + U$. As discussed in Carroll et al. (1999), for any fixed $\lambda > 0$, after repeatedly adding the extra normal measurement errors, and computing the ordinary least squares (LS) slope, the averaged estimator consistently estimates $g(\lambda) = \beta\sigma_x^2/(\sigma_x^2+(1+\lambda)\sigma_u^2)$, where σ_x^2 is the variance of X. Obviously, extrapolating λ to -1 leads to $g(-1) = \beta$. This clearly shows that the SIMEX method works really well for the linear regression model. In fact, in the seminal paper of Cook and Stefanski (1994), the SIMEX estimators of α and β are derived without the simulation step. Instead, the conditional expectation of the least square estimates based on the pseudo-data given the observed data are calculated, and under the NON-IID pseudo-errors, the following estimators of α and β are constructed:

$$\hat{\alpha}(\lambda) = \overline{Y} - \hat{\beta}(\lambda)\overline{Z}, \quad \hat{\beta}(\lambda) = \frac{S_{YZ}}{S_{ZZ} + \lambda\sigma_u^2}, \tag{2.2.1}$$

where $\overline{Y}, \overline{X}$ are the sample means of Y and Z, S_{YZ}, S_{ZZ} are the sample covariance of Y and Z, the sample variance of Z, respectively. Directly letting $\lambda = -1$ leads to the SIMEX estimator. For more details regarding the NON-IID pseudo-errors, please refer to Cook and Stefanski (1994) and Section 5.3.4.1 in Carroll et al. (2006). In the following, we would like to show that the SIMEX estimators of α, β defined in (2.2.1) can be obtained from another perspective, without using the so-called NON-IID pseudo-errors.

Consider a multiple linear regression model and the LS procedure. When X is observable, the LS estimators of α and β can be estimated by minimizing the LS criterion $\sum_{i=1}^{n}(Y_i - \alpha - \beta^T X_i)^2$. Since X_i's are not available, following the SIMEX idea, we generate the pseudo-data $Z_i(\lambda) = Z_i + \sqrt{\lambda} V_i$, $V_i \sim N(0, \Sigma_u)$, $i = 1, 2, \ldots, n$, where Σ_u denotes the covariance matrix of U which is assumed to be known. However, before minimizing $\sum_{i=1}^{n}(Y_i - \alpha - \beta^T Z_i(\lambda))^2$ and following the classcial SIMEX road map, we minimize the following conditional expectation

$$E\left[\sum_{i=1}^{n}(Y_i - \alpha - \beta^T Z_i(\lambda))^2 | (\boldsymbol{Y}, \boldsymbol{Z})\right], \tag{2.2.2}$$

where $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ and $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$. Since V_i's are i.i.d. from $N(0, \Sigma_u)$ and independent of other random variables in the model, so the expectation (2.2.2) equals

$$\sum_{i=1}^{n}(Y_i - \alpha - \beta^T Z_i)^2 + n\lambda \boldsymbol{\beta}^T \Sigma_U \boldsymbol{\beta}.$$

The minimizer of the above expression is simply

$$\hat{\beta}(\lambda) = (S_{ZZ} + \lambda \Sigma_U)^{-1} S_{YZ}, \quad \hat{\alpha}(\lambda) = \overline{Y} - \hat{\beta}^T(\lambda)\overline{X}, \tag{2.2.3}$$

and by choosing $\lambda = -1$, we immediately have the commonly used bias-corrected estimators or the SIMEX estimators derived using the NON-IID pseudo-errors. Note that here not only do we not need the simulation step, but also the extrapolation step is unnecessary.

Encouraged by this simple example, we dive into some more complex regression models to see if the above method can be applied to more general regression setups. In the following we use $m(x)$ to denote a regression function, θ_0 the true value of θ, and the regression error is independent of the covariate X.

Example 1: Suppose $m(x) = \exp(x^T\theta)$. Then

$$E[(Y - m(Z(\lambda), \theta))^2|Y, Z] = E[(Y - \exp(Z^T\theta + \sqrt{\lambda}U^T\theta))^2|Y, Z]$$
$$= Y^2 - 2Y\exp(Z^T\theta)E\exp(\sqrt{\lambda}U^T\theta) + \exp(2Z^T\theta)E\exp(2\sqrt{\lambda}U^T\theta)$$
$$= Y^2 - 2Y\exp(Z^T\theta)\exp(\lambda\theta^T\Sigma_u\theta/2) + \exp(2Z^T\theta)\exp(2\lambda\theta^T\Sigma_u\theta).$$

Therefore, we can estimate θ by the minimizer of the empirical version of the above conditional expectation with λ replaced with -1,

$$L_n(\theta) := \frac{1}{n}\sum_{i=1}^{n}\left[Y_i^2 - 2Y_i\exp(Z_i^T\theta - \theta^T\Sigma_u\theta/2) + \exp(2Z_i^T\theta - 2\theta^T\Sigma_u\theta)\right].$$

To see why minimizing the above target function leads to a consistent estimator, we take the derivative of $L_n(\theta)$ with respect to θ and set it to 0. That is

$$\frac{1}{n}\sum_{i=1}^{n}Y_i(Z_i - \Sigma_u\theta)\exp(Z_i^T\theta) - \frac{1}{n}\sum_{i=1}^{n}(Z_i - 2\Sigma_u\theta)\exp(2Z_i^T\theta - 3\theta^T\Sigma_u\theta/2) = 0. \qquad (2.2.4)$$

Note that, as $n \to \infty$, almost surely,

$$\frac{1}{n}\sum_{i=1}^{n}Y_iZ_i\exp(Z_i^T\theta) \to EYZ\exp(Z^T\theta) = \exp(\theta^T\Sigma_u\theta/2)E\left[(X + \Sigma_u\theta)\exp\left((\theta+\theta_0)^TX\right)\right],$$
$$\frac{1}{n}\sum_{i=1}^{n}Y_i\exp(Z_i^T\theta) \to \exp(\theta^T\Sigma_u\theta/2)E\exp\left((\theta+\theta_0)^TX\right),$$
$$\frac{1}{n}\sum_{i=1}^{n}\exp(2Z_i^T\theta) \to E\exp\left(2\theta^TX\right)\exp(2\theta^T\Sigma_u\theta),$$
$$\frac{1}{n}\sum_{i=1}^{n}Z_i\exp(2Z_i^T\theta) \to E\left[(X + 2\Sigma_u\theta)\exp\left(2\theta^TX\right)\right]\exp(2\theta^T\Sigma_u\theta).$$

Therefore, as $n \to \infty$, almost surely,

$$\dot{L}_n(\theta) \to$$

$$\exp(\theta^T \Sigma_u \theta/2) E(X + \Sigma_u \theta) \exp\left((\theta+\theta_0)^T X\right) - \exp(\theta^T \Sigma_u \theta/2) E \exp\left((\theta+\theta_0)^T X\right) \Sigma_u \theta$$
$$- \exp(\theta^T \Sigma_u \theta/2) E(X + 2\Sigma_u \theta) \exp\left(2\theta^T X\right) + 2\exp(\theta^T \Sigma_u \theta/2) E \exp\left(2\theta^T X\right) \Sigma_u \theta.$$

Denote the limit as $\dot{L}(\theta)$. Then $\dot{L}(\theta) = 0$ if and only if

$$EX \exp\left((\theta+\theta_0)^T X\right) - EX \exp\left(2\theta^T X\right) = 0.$$

Obviously, θ_0 is a solution.

Example 2: Suppose $m(x) = \sin(x^T \theta)$. Then

$$\begin{aligned}
&E[(Y - m(Z(\lambda),\theta))^2 | Y, Z] = E[(Y - \sin(Z^T\theta + \sqrt{\lambda} U^T \theta))^2 | Y, Z] \\
&= Y^2 - 2Y \sin(Z^T\theta) E \cos(\sqrt{\lambda} U^T\theta) - 2Y \cos(Z^T\theta) E \sin(\sqrt{\lambda} U^T\theta) \\
&\quad + \sin^2(Z^T\theta) E \cos^2(\sqrt{\lambda} U^T\theta) + \cos^2(Z^T\theta) E \sin^2(\sqrt{\lambda} U^T\theta) \\
&\quad + 2 \sin(Z^T\theta) \cos(Z^T\theta) E \cos(\sqrt{\lambda} U^T\theta) \sin(\sqrt{\lambda} U^T\theta) \\
&= Y^2 - 2Y \sin(Z^T\theta) E \cos(\sqrt{\lambda} U^T\theta) \\
&\quad + \sin^2(Z^T\theta) E \cos^2(\sqrt{\lambda} U^T\theta) + \cos^2(Z^T\theta) E \sin^2(\sqrt{\lambda} U^T\theta).
\end{aligned}$$

Note that

$$E \cos(\sqrt{\lambda} U^T\theta) = E \exp\left(\mathbf{i} \sqrt{\lambda} U^T \theta\right) = \exp\left(-\frac{1}{2}\lambda \theta^T \Sigma_u \theta\right),$$

and from the well known trigonometric identities $\sin^2 u = (1 - \cos(2u))/2$ and $\cos^2 u = (1 + \cos(2u))/2$, we have

$$E \cos^2(\sqrt{\lambda} U^T\theta) = \frac{1}{2} + \frac{1}{2} E \cos(2\sqrt{\lambda} U^T\theta) = \frac{1}{2} + \frac{1}{2} \exp\left(-2\lambda \theta^T \Sigma_u \theta\right),$$

$$E\sin^2(\sqrt{\lambda}U^T\theta) = \frac{1}{2} - \frac{1}{2}E\cos(2\sqrt{\lambda}U^T\theta) = \frac{1}{2} - \frac{1}{2}\exp\left(-2\lambda\theta^T\Sigma_u\theta\right).$$

Therefore, by $\cos^2 u - \sin^2 u = \cos(2u)$, we get

$$\begin{aligned} &E[(Y - m(Z(\lambda),\theta))^2|Y,Z] \\ = &Y^2 - 2Y\sin(Z^T\theta)\exp\left(-\frac{1}{2}\lambda\theta^T\Sigma_u\theta\right) - \frac{1}{2}\cos(2Z^T\theta)\exp\left(-2\lambda\theta^T\Sigma_u\theta\right) + 1, \end{aligned}$$

and an estimator of θ can be obtained by minimizing

$$\frac{1}{n}\sum_{i=1}^n\left[Y_i^2 - 2Y_i\sin(Z_i^T\theta)\exp\left(\frac{1}{2}\theta^T\Sigma_u\theta\right) - \frac{1}{2}\cos(2Z_i^T\theta)\exp\left(2\theta^T\Sigma_u\theta\right)\right] + 1.$$

By a simple algebra, we can show that as $n \to \infty$, the above average converges almost surely to $1/2 + \sigma_\varepsilon^2 + E[\sin(X^T\theta_0) - \sin(X^T\theta)]^2$, which is minimized at $\theta = \theta_0$.

Example 3 (Poisson Regression): Given a p-dimensional covariate X, suppose a non-negative integer-valued random variable Y has a Poisson distribution with mean $\exp(X^T\theta)$. If a sample $(X_i, Y_i), i = 1, 2, \ldots, n$ can be drawn from (X, Y), then we can estimate θ by maximizing the log-likelihood function $\sum_{i=1}^n[Y_iX_i^T\theta - \exp(X_i^T\theta) - \log Y_i!]$. Note that

$$\begin{aligned} E[YZ^T(\lambda)\theta - \exp(Z^T(\lambda)\theta)|(Y,Z)] &= YZ^T\theta - \exp(Z^T\theta)E(\exp(\sqrt{\lambda}V^T\theta)) \\ &= YZ^T\theta - \exp(Z^T\theta)\exp(\lambda\theta^T\Sigma_u\theta/2). \end{aligned}$$

Thus, by directly extrapolating λ to -1, we can estimate θ using the maximizer of

$$\sum_{i=1}^n[Y_iZ_i^T\theta - \exp(Z_i^T\theta - \theta^T\Sigma_u\theta/2)]$$

which coincides with the estimation procedure proposed in Guo and Li (2002).

The three examples discussed above demonstrate that in some cases, $\lambda = -1$ can be directly plugged into the conditional expectations and an estimator of θ can be obtained by solving a standard nonlinear equation system. However, this nice property is not shared by other models universally.

Example 4 (Logistic Regression): In logistic regression, a 0-1 random variable Y depends on a p-dimensional covariate X via the probability distribution $P(Y = 1) = 1 - P(Y = 0) = F(\alpha + X^T\beta)$, where $F(x) = 1/(1 + \exp(-x))$ is the logistic function. The log-likelihood function of α, β based on a sample (Y_i, X_i), $i = 1, 2, \ldots, n$ is given by $\sum_{i=1}^n \left\{Y_i(\alpha + X_i^T\beta) - \log[1 + \exp(\alpha + X_i^T\beta)]\right\}$. With $Z(\lambda) = Z + \sqrt{\lambda}V$, we have

$$\begin{aligned} & E[Y(\alpha + Z^T(\lambda)\beta)|(Y, Z)] - E\left\{\log[1 + \exp(\alpha + Z^T(\lambda)\beta)]|(Y, Z)\right\} \\ = & Y(\alpha + Z^T\beta) - \int \log[1 + \exp(\alpha + Z^T\beta + u)]\phi(u, 0; \lambda\beta^T\Sigma_u\beta)du. \end{aligned}$$

Note that $\lambda = -1$ can not be plugged into the integration.

Example 5: Consider the multiplicative regression model $Y = \exp(X^T\theta)\varepsilon$, where ε is a positive random variable, with $E(\varepsilon|X) = 1$. This model is often called the accelerated failure time (AFT) model and is widely used in the survival analysis, econometrics and finance areas. There are two methods in the literature to estimate θ when observations can be made on (Y, X). The first one is the least absolute relative error (LARE) estimation procedure which minimizes the following criterion

$$LARE(\theta) = \sum_{i=1}^n \left[Y_i^{-1}|Y_i - \exp(X_i^T\theta)| + \exp(-X_i^T\theta)|Y_i - \exp(X_i^T\theta)|\right]$$

proposed by Chen et al. (2010). The second method is called the least product relative error (LPRE) estimation which minimizes the following expression

$$LPRE(\theta) = \sum_{i=1}^n \left[Y_i \exp(-X_i^T\theta) + Y_i^{-1}\exp(X_i^T\theta)\right]$$

proposed in Chen et al. (2016).

In the case of X being contaminated with normal measurement errors, according to the previous arguments, for the LARE criterion, we may consider the following conditional expectation:

$$E\left\{\left[Y^{-1}|Y-\exp(Z^T(\lambda)\theta)|+\exp(-Z^T(\lambda)\theta)|Y-\exp(Z^T(\lambda)\theta)|\right]\Big|(Y,Z)\right\} \tag{2.2.5}$$

and for the LPRE criterion, we will calculate

$$E\left\{\left[Y\exp(-Z^T(\lambda)\theta)+Y^{-1}\exp(Z^T(\lambda)\theta)\right]\Big|(Y,Z)\right\}. \tag{2.2.6}$$

For (2.2.5), we have

$$\begin{aligned}
&E\left\{Y^{-1}|Y-\exp(Z^T(\lambda)\theta)|\Big|(Y,Z)\right\}=\int\left\{Y^{-1}|Y-\exp(Z^T\theta+v)|\right\}\phi(v;0,\lambda\theta^T\Sigma_u\theta)dv\\
&=2\Phi(\log(Y\exp(-Z^T\theta));0,\lambda\theta^T\Sigma_u\theta)-1\\
&\quad+Y^{-1}\exp(Z^T\theta+\lambda\theta^T\Sigma_u\theta/2)\left[1-2\Phi(\log(Y\exp(-Z^T\theta))+\lambda\theta^T\Sigma_u\theta,0,\lambda\theta^T\Sigma_u\theta)\right],
\end{aligned}$$

and

$$\begin{aligned}
&E\left\{\exp(-Z^T(\lambda)\theta)|Y-\exp(Z^T(\lambda)\theta)|\Big|(Y,Z)\right\}\\
&=\int\left\{\exp(-Z^T\theta+v)|Y-\exp(Z^T\theta+v)|\right\}\phi(v;0,\lambda\theta^T\Sigma_u\theta)dv\\
&=1-2\Phi(\log(Y\exp(-Z^T\theta));0,\lambda\theta^T\Sigma_u\theta)\\
&\quad+Y\exp(-Z^T\theta+\lambda\theta^T\Sigma_u\theta/2)\left[2\Phi(\log(Y\exp(-Z^T\theta))+\lambda\theta^T\Sigma_u\theta,0,\lambda\theta^T\Sigma_u\theta)-1\right].
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&E\left\{\left[Y^{-1}|Y-\exp(Z^T(\lambda)\theta)|+\exp(-Z^T(\lambda)\theta)|Y-\exp(Z^T(\lambda)\theta)|\right]\Big|(Y,Z)\right\}\\
&=Y^{-1}\exp(Z^T\theta+\lambda\theta^T\Sigma_u\theta/2)\left[1-2\Phi(\log(Y\exp(-Z^T\theta))+\lambda\theta^T\Sigma_u\theta,0,\lambda\theta^T\Sigma_u\theta)\right]\\
&\quad+Y\exp(-Z^T\theta+\lambda\theta^T\Sigma_u\theta/2)\left[2\Phi(\log(Y\exp(-Z^T\theta))+\lambda\theta^T\Sigma_u\theta,0,\lambda\theta^T\Sigma_u\theta)-1\right].
\end{aligned}$$

Note that we cannot directly plug $\lambda = -1$ into the above expectation, since $\lambda\theta^T\Sigma_u\theta$ serves as the variance of a normal distribution. However, for (2.2.6), we have

$$
\begin{aligned}
&E\left[Y\exp(-Z^T(\lambda)\theta) + Y^{-1}\exp(Z^T(\lambda)\theta)\Big|(Y,Z)\right] \\
&= \left[Y\exp(-Z^T\theta) + Y^{-1}\exp(Z^T\theta)\right]\exp\left(\lambda\theta^T\Sigma_u\theta/2\right).
\end{aligned}
$$

$\lambda = -1$ can be directly plugged in.

Example 6 (Quantile Regression): For a positive number $\tau \in (0,1)$, let $\rho_\tau(x) = x(\tau - I(x < 0))$. Then the quantile regression estimates the regression coefficients β by the minimizer of $\sum_{i=1}^n \rho_\tau(Y_i - X_i^T\beta)$ when both Y and X are observable. In the measurement error setup, similar to the previous examples, we may estimate β by maximizing the conditional expectation $\sum_{i=1}^n E[\rho_\tau(Y_i - Z_i^T(\lambda)\beta)|(Y_i, Z_i)]$. Calculation shows that

$$
\begin{aligned}
&E[\rho_\tau(Y - Z^T(\lambda)\beta)|(Y,Z)] \\
&= E\left[(Y - Z^T\beta - \sqrt{\lambda}V^T\beta)(\tau - I(Y - Z^T\beta - \sqrt{\lambda}V^T\beta < 0))|(Y,Z)\right] \\
&= \tau(Y - Z^T\beta) - (Y - Z^T\beta)\int_{Y-Z^T\beta}^{\infty}\phi(v; 0, \lambda\beta^T\Sigma_u\beta)dv + \int_{Y-Z^T\beta}^{\infty} v\phi(v; 0, \lambda\beta^T\Sigma_u\beta)dv \\
&= (\tau - 1)(Y - Z^T\beta) + (Y - Z^T\beta)\Phi(Y - Z^T\beta; 0, \lambda\beta^T\Sigma_u\beta) + \lambda\beta^T\Sigma_u\beta\phi(Y - Z^T\beta; 0, \lambda\beta^T\Sigma_u\beta)].
\end{aligned}
$$

Denote $\xi_i(\beta) = Y_i - Z_i^T\beta$. We can see that the target function has the form of

$$
(\tau - 1)\sum_{i=1}^n \xi_i(\beta) + \sum_{i=1}^n \xi_i(\beta)\Phi(\xi_i(\beta); 0, \lambda\beta^T\Sigma_u\beta) + \lambda\beta^T\Sigma_u\beta\sum_{i=1}^n \phi(\xi_i(\beta); 0, \lambda\beta^T\Sigma_u\beta).
$$

It is easy to see that the new target function is a nonlinear differentiable function of β, and it can be readily minimized using standard algorithms.

Example 7 (Walsh Regression): Another robust estimation procedure is the Walsh-average regression proposed in Feng et al. (2012). For a response variable Y and a covariate vector X, the Walsh-average regression estimates the regression coefficient β by minimizing the following objective function,

$$\frac{1}{2n(n+1)}\sum_{i\le j}|Y_i - X_i^T\beta + Y_j - X_j^T\beta| := \frac{1}{2n(n+1)}\sum_{i\le j} L(Y_i, Y_j, X_i, X_j; \beta)$$

based on a sample $(Y_i, X_i), i = 1, \ldots, n$ from (Y, X). When X_i's are measured with normal measurement error U_i, then similar to the previous arguments, we may estimate β by maximizing the conditional expectation

$$\begin{aligned}&\sum_{i\le j} E[L(Y_i, Y_j, Z_i^T(\lambda), Z_i^T(\lambda); \beta)|(Y_i, Y_j, Z_i, Z_j)]\\ &= 2\sum_{i=1}^n E[|Y_i - Z_i^T(\lambda)\beta||(Y_i, Z_i)] + \sum_{i<j} E[|Y_i + Y_j - (Z_i + Z_j)^T(\lambda)\beta||(Y_i, Y_j, Z_i, Z_j)].\end{aligned}$$

Based on the discussion on the quantile regression, taking $\tau = 1/2$, the first term on the right-hand side equals

$$2\sum_{i=1}^n \xi_i(\beta)[2\Phi(\xi_i(\beta), 0, \lambda\beta^T\Sigma_u\beta) - 1] + 4\lambda\beta^T\Sigma_u\beta\sum_{i=1}^n \phi(\xi_i(\beta), 0, \lambda\beta^T\Sigma_u\beta),$$

and the second term on the right-side equals

$$\sum_{i<j} \xi_{ij}(\beta)[2\Phi(\xi_{ij}(\beta), 0, 2\lambda\beta^T\Sigma_u\beta) - 1] + 8\lambda\beta^T\Sigma_u\beta\sum_{i<j} \phi(\xi_{ij}(\beta), 0, 2\lambda\beta^T\Sigma_u\beta),$$

where $\xi_{ij}(\beta) = Y_i + Y_j - (Z_i + Z_j)^T\beta$. Again, this leads to a nonlinear target function to implement the proposed estimation procedure.

Example 8 (Expectile Regression): For a positive number $\tau \in (0, 1)$, let $\rho_\tau(x) = x^2[\tau - I(x < 0)]$. Then the expectile regression estimates the parameter β by minimizing the target function $\sum_{i=1}^n \rho_\tau(Y_i - X_i^T\beta)$ when both Y and X are observable. When X is contaminated with normal measurement error, β might be estimated by the maximizer of the conditional expectation $\sum_{i=1}^n E[\rho_\tau(Y_i - Z_i^T(\lambda)\beta)|(Y_i, Z_i)]$. Calculation shows that

$$
\begin{aligned}
& E[\rho_\tau(Y - Z^T(\lambda)\beta)|(Y, Z)] \\
&= E\left[(Y - Z^T\beta - \sqrt{\lambda}V^T\beta)^2[\tau - I(Y - Z^T\beta - \sqrt{\lambda}V^T\beta < 0)]|(Y, Z)\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \int (Y - Z^T\beta - v)^2|\tau - I(Y - Z^T\beta - v < 0)|\phi(v; 0, \lambda\beta^T\Sigma_u\beta)dv \\
&= \tau \int_{-\infty}^{Y - Z^T\beta} [(Y - Z^T\beta)^2 - 2v(Y - Z^T\beta) + v^2]\phi(v; 0, \lambda\beta^T\Sigma_u\beta)dv \\
&\quad +(1-\tau)\int_{Y - Z^T\beta}^{\infty} [(Y - Z^T\beta)^2 - 2v(Y - Z^T\beta) + v^2]\phi(v; 0, \lambda\beta^T\Sigma_u\beta)dv \\
&= \tau(Y - Z^T\beta)^2\Phi(Y - Z^T\beta; 0, \lambda\beta^T\Sigma_u\beta) - 2\tau(Y - Z^T\beta)\int_{-\infty}^{Y - Z^T\beta} v\phi(v; 0, \lambda\beta^T\Sigma_u\beta)dv \\
&\quad +\tau\int_{-\infty}^{Y - Z^T\beta} v^2\phi(v; 0, \lambda\beta^T\Sigma_u\beta)dv + (1-\tau)(Y - Z^T\beta)^2[1 - \Phi(Y - Z^T\beta; 0, \lambda\beta^T\Sigma_u\beta)] \\
&\quad -2(1-\tau)(Y - Z^T\beta)\int_{Y - Z^T\beta}^{\infty} v\phi(v; 0, \lambda\beta^T\Sigma_u\beta)dv + (1-\tau)\int_{Y - Z^T\beta}^{\infty} v^2\phi(v; 0, \lambda\beta^T\Sigma_u\beta)dv \\
&= \tau(Y - Z^T\beta)^2\Phi(Y - Z^T\beta; 0, \lambda\beta^T\Sigma_u\beta) + 2\tau\lambda\beta^T\Sigma_u\beta(Y - Z^T\beta)\phi(Y - Z^T\beta; 0, \lambda\beta^T\Sigma_u\beta) \\
&\quad -\tau\lambda\beta^T\Sigma_u\beta[(Y - Z^T\beta)\phi(v; 0, \lambda\beta^T\Sigma_u\beta) - \Phi(v; 0, \lambda\beta^T\Sigma_u\beta)] \\
&\quad +(1-\tau)(Y - Z^T\beta)^2[1 - \Phi(Y - Z^T\beta; 0, \lambda\beta^T\Sigma_u\beta)] \\
&\quad -2(1-\tau)\lambda\beta^T\Sigma_u\beta(Y - Z^T\beta)\phi(Y - Z^T\beta; 0, \lambda\beta^T\Sigma_u\beta) \\
&\quad +(1-\tau)\lambda\beta^T\Sigma_u\beta[(Y - Z^T\beta)\phi(Y - Z^T\beta; 0, \lambda\beta^T\Sigma_u\beta) + 1 - \Phi(Y - Z^T\beta; 0, \lambda\beta^T\Sigma_u\beta)] \\
&= (2\tau - 1)(Y - Z^T\beta)^2\Phi(Y - Z^T\beta; 0, \lambda\beta^T\Sigma_u\beta) \\
&\quad +(2\tau - 1)\lambda\beta^T\Sigma_u\beta(Y - Z^T\beta)\phi(Y - Z^T\beta; 0, \lambda\beta^T\Sigma_u\beta) \\
&\quad +(2\tau - 1)\lambda\beta^T\Sigma_u\beta\Phi(Y - Z^T\beta; 0, \lambda\beta^T\Sigma_u\beta) + (1-\tau)[(Y - Z^T\beta)^2 + \lambda\beta^T\Sigma_u\beta].
\end{aligned}
$$

Recall the notations $\xi_i(\beta)$, we can see that the target function takes the form of

$$(2\tau-1)\sum_{i=1}^{n}\xi_i^2(\beta)\Phi(\xi_i(\beta);0,\lambda\beta^T\Sigma_u\beta)+(2\tau-1)\lambda\beta^T\Sigma_u\beta\sum_{i=1}^{n}\xi_i(\beta)\phi(\xi_i(\beta);0,\lambda\beta^T\Sigma_u\beta)$$
$$+(2\tau-1)\lambda\beta^T\Sigma_u\beta\sum_{i=1}^{n}\Phi(\xi_i(\beta);0,\lambda\beta^T\Sigma_u\beta)+(1-\tau)\sum_{i=1}^{n}[\xi_i^2(\beta)+\lambda\beta^T\Sigma_u\beta].$$

Like the quantile regression in Example 6, the new target function for the expectile regression is also a nonlinear differentiable function of β, and it can be readily minimized using standard algorithms.

Most estimation procedures are built upon optimizing specific target functions or searching for the solution of certain estimating equations. The above interesting findings may suggest that, after replacing the true predictors with the pseudo-data in the target functions or the estimating equations, one can simply optimize the conditional expectation or solve the new estimating equations. This allows researchers to circumvent the computationally-intensive simulation step of the classical SIMEX procedure. Also, if the process goes smoothly, the conditional expectation can be directly extrapolated to $\lambda=-1$ and an estimate can be obtained by solving a standard nonlinear equation, as shown in Examples 1, 2, 3, as well as the LPRE procedure in Example 4. In cases such as those described in Examples 4-8, where directly plugging $\lambda=-1$ into the resulting expectation is not feasible, one can proceed with the extrapolation step to obtain an estimate for θ. In Section 2.3, we shall formulate the extrapolation estimation procedure for the general parametric regression models, and discuss its statistical properties.

2.3 Extrapolation Estimation in Parametric Regression

For a general parametric regression model $Y=m(X;\theta)+\varepsilon$ with $Z=X+U$, where $X\in\mathbb{R}^p$, ε, X, U are independent, $\theta\in\Theta\subset\mathbb{R}^q$, and p,q are some positive integers, we may have different ways to estimate θ based on various assumptions on the model. In this section, the least squares estimation (LSE) procedure will be used as an exemplary method to construct the extrapolation estimation procedure. That is, the following conditional expectation

$$E\left[\sum_{i=1}^{n}(Y_i - m(Z_i(\lambda);\theta))^2|(\boldsymbol{Y},\boldsymbol{Z})\right] = \sum_{i=1}^{n}\int(Y_i - m(Z_i+u;\theta))^2\phi(u,0,\lambda\Sigma_u)du$$

with respect to θ will be minimized. To see intuitively why extrapolating λ to -1 can result in a reasonable estimate of θ_0, the true value of θ, in this general setup as the sample size $n \to \infty$, we denote $\theta(\lambda) = \text{argmin}_\theta L(\theta;\lambda)$, where $L(\theta;\lambda) = E\int(Y - m(Z+u;\theta))^2\phi(u,0,\lambda\Sigma_u)du$, and

$$L_n(\theta;\lambda) = \frac{1}{n}\sum_{i=1}^{n}\int(Y_i - m(Z_i+u;\theta))^2\phi(u,0,\lambda\Sigma_u)du. \tag{2.3.1}$$

Under some regularity conditions, by the strong law of large numbers, $L_n(\theta;\lambda) \Longrightarrow L(\theta;\lambda)$ almost surely as $n \to \infty$. This, together with the fact

$$L(\theta;\lambda) = E\int(Y - m(X+u;\theta))^2\phi(u,0,(\lambda+1)\Sigma_u)du \Longrightarrow E(Y - m(X;\theta))^2 \tag{2.3.2}$$

as $\lambda \to -1$ if we assume that $m(x;\theta)$ is continuous in x for each $\theta \in \Theta$, implies that $\theta(\lambda) \to \theta_0$ if the equation $E(m(X;\theta_0) - m(X;\theta))^2 = 0$ has a unique solution. This heuristic argument leads to the following extrapolation algorithm for estimating θ:

Extrapolation Estimation Algorithm

1. If $L_n(\theta;\lambda)$ can be directly extrapolated to $\lambda = -1$, then an estimate of θ_0 is given by $\hat{\theta}_n = \text{argmin}_\theta L_n(\theta; -1)$, where $L_n(\theta;\lambda)$ is defined in (2.3.1).

2. If $L_n(\theta;\lambda)$ cannot be directly extrapolated to $\lambda = -1$, then pre-select some grid points $0 = \lambda_1 < \lambda_2 < \cdots < \lambda_K = c$, for example, $c = 2$.

   - For each $\lambda \in \{\lambda_1, \ldots, \lambda_K\}$, solving $\hat{\theta}_n(\lambda) = \text{argmin}_\theta L_n(\theta;\lambda)$.
   - Fit a trend for the pairs $(\lambda_k, \hat{\theta}_n(\lambda_k))$, $k = 1, 2, \ldots, K$, and extrapolate this trend back to -1.

   Take the extrapolated value $\hat{\theta}_n$ as an estimate of θ_0.

In the following we shall explore the extrapolation estimation algorithm described above in detail by presenting some theoretical results. For a generic parametric function $g(x;\theta)$ with multidimensional x and θ, we denote $\dot{g}(x;\theta) = \partial g(x;\theta)/\partial\theta$, $\ddot{g}(x;\theta) = \partial^2 g(x;\theta)/\partial\theta\partial\theta^T$, and $g'(x;\theta) = \partial g(x;\theta)/\partial x$, $g''(x;\theta) = \partial^2 g(x;\theta)/\partial x \partial x^T$.

First we list some technical assumptions needed for presenting the theoretical arguments.

(C1). The parameter space Θ of θ is compact;

(C2). For each $x \in \mathbb{R}^p$, the regression function $m(x;\theta)$ is twice continuously differentiable for each $\theta \in \Theta$;

(C3). There exists a function $K(x)$, not depending on θ, such that $EK^2(Z+V) < \infty$ and $|\ddot{m}(x;\theta)| + |m(x;\theta)| \leq K(x)$ for all x in the domain of m;

(C4). For each $\lambda \geq 0$, the minimizer of $L(\theta, \lambda)$ exists and is unique;

(C5). $\text{argmin}_{\theta\in\Theta}\, E[m(X;\theta) - m(X;\theta_0)]^2$ is unique;

(C6). For each $\lambda \geq 0$, $E\left[\dot{m}(Z+V;\theta)\dot{m}^T(Z+V;\theta)\right]$ is positive definite, where $V \sim N(0, \lambda\Sigma_u)$, Z and V are independent.

Conditions(C1)-(C3) allow us to show the uniform convergence of $L_n(\theta)$ to $L(\theta)$ over $\theta \in \Theta$ using the uniform convergence theory discussed in Ferguson (1996), which, together with the condition (C4), implies the convergence of $\hat{\theta}_n(\lambda)$ to $\theta(\lambda)$ as $n \to \infty$ for each $\lambda > 0$, and eventually to θ_0 by letting $\lambda \to -1$ by using (C5). In fact, if $\text{argmin}_\theta L(\theta; \lambda)$ is not unique, then we can show that for any local minimizer, there is a sequence of minimizers of $L_n(\theta; \lambda)$ that converges to the local minimizer in probability. However, to keep the argument relatively simple, we shall adopt (C5) in the following discussion. Finally, using (C6), we can show the asymptotic normality of the extrapolation estimator. The following theorem summarizes the consistency of $\hat{\theta}_n(\lambda)$ to $\theta(\lambda)$ for each $\lambda > 0$, and the approximation of $\theta(\lambda)$ to θ_0 as λ is extrapolated to -1.

Theorem 1. *Suppose that the conditions (C1), (C2), (C3), and (C4) hold. Then for each* $\lambda > 0$, $\hat{\theta}_n(\lambda) \to \theta(\lambda)$ *in probability as* $n \to \infty$. *If we further assume that (C5) holds, then as* $\lambda \to -1$,

$$\theta(\lambda) = \theta_0 - (\lambda + 1)\left[E\dot{m}(X;\theta_0)\dot{m}^T(X;\theta_0)\right]^{-1} \cdot \\ \left[E\dot{m}(X;\theta_0)trace(m''(X;\theta_0)\Sigma_u^2) + E\dot{m}'(X;\theta_0)\Sigma_u^2 m'(X;\theta_0)\right] + o((\lambda+1)).$$

Denote $\Lambda = (\lambda_1, \ldots, \lambda_K)^T$, $\hat{\theta}_n(\Lambda) = (\hat{\theta}_n(\lambda_1), \ldots, \hat{\theta}_n(\lambda_K))$, $\theta(\Lambda) = (\theta(\lambda_1), \ldots, \theta(\lambda_K))$. The asymptotic joint normality of $\hat{\theta}_n(\Lambda)$ to $\theta(\Lambda)$ is described in the following theorem.

Theorem 2. *In addition to the conditions in Theorem 1, suppose (C6) holds. Then we have*

$$\sqrt{n}[(\hat{\theta}_n(\lambda_1) - \theta(\lambda_1))^T, \ldots, (\hat{\theta}_n(\lambda_K) - \theta(\lambda_K))^T]^T \sim N(0, \Omega_1^{-1}(\Lambda)\Omega_0(\Lambda)\Omega_1^{-1}(\Lambda)),$$

where

$$\Omega_0(\Lambda) = [\Sigma_0(\lambda_j, \lambda_k)]_{K\times K}, \quad \Omega_1(\Lambda) = Diag(\Sigma_1(\lambda_1), \ldots, \Sigma_1(\lambda_K)),$$
$$\Sigma_1(\lambda) = E\int \dot{m}(X+u;\theta(\lambda))\dot{m}^T(X+u;\theta(\lambda))\phi(u, 0, (\lambda+1)\Sigma_u)du \\ -E\int [m(X;\theta_0) - m(X+u, \theta(\lambda))]\ddot{m}(X+u;\theta(\lambda))\phi(u, 0, (\lambda+1)\Sigma_u)du$$

and $\Sigma_0(\lambda_j, \lambda_l)$*,* $j, l = 1, 2, \ldots, K$ *is defined as*

$$E\iint \left\{\sigma_\varepsilon^2 + [m(X, \theta_0) - m(X + u, \theta(\lambda_j))][m(X, \theta_0) - m(X + v, \theta(\lambda_l))]\right\} \dot{m}(X + u; \theta(\lambda_j))$$
$$\dot{m}^T(X + v; \theta(\lambda_l))\phi(u, v, (\lambda_j + \lambda_l)\Sigma_u)\phi\left(\frac{\lambda_j v + \lambda_l u}{\lambda_j + \lambda_l}, 0, \frac{\lambda_j + \lambda_l + \lambda_j\lambda_l}{\lambda_j + \lambda_l}\Sigma_u\right) dudv$$

or

$$E\iint \left\{\sigma_\varepsilon^2 + [m(X, \theta_0) - m(X + u, \theta(\lambda_j))][m(X, \theta_0) - m(X + v, \theta(\lambda_l))]\right\} \dot{m}(X + u; \theta(\lambda_j))$$
$$\dot{m}^T(X + v; \theta(\lambda_l))\phi(v, 0, (\lambda_l + 1)\Sigma_u)\phi\left(u, \frac{v}{\lambda_l + 1}, \frac{\lambda_1 + \lambda_l + \lambda_j\lambda_l}{\lambda_l + 1}\Sigma_u\right) dudv.$$

From Theorem 2, one can see that for each $\lambda > 0$,

$$\sqrt{n}(\hat{\theta}_n(\lambda) - \theta(\lambda)) \to N\left(0, \Sigma_1^{-1}(\lambda)\Sigma_0(\lambda, \lambda)\Sigma_1^{-1}(\lambda)\right)$$

in distribution. In particular, for $\lambda = 0$, we have

$$\sqrt{n}(\hat{\theta}_n(0) - \theta(0)) \to N(0, \Sigma_1^{-1}(0)\Sigma_0(0, 0)\Sigma_1^{-1}(0)),$$

where $\theta(0) = \text{argmin}_{\theta\in\Theta} E[m(X; \theta_0) - m(Z; \theta)]^2$, and

$$\Sigma_0(0, 0) = E\left\{\sigma_\varepsilon^2 + [m(X, \theta_0) - m(Z, \theta(0))][m(X, \theta_0) - m(Z, \theta(0))]\right\} \dot{m}(Z; \theta(0))\dot{m}^T(Z; \theta(0)),$$
$$\Sigma_1(0) = E\dot{m}(Z; \theta(0))\dot{m}^T(Z; \theta(0)) - E[m(X; \theta_0) - m(Z, \theta(0))]\ddot{m}(Z; \theta(0)).$$

This recovers the asymptotic normality result for the naive estimator $\hat{\theta}_n(0)$ of θ_0. It is in our interest to know the limits of $\Sigma_1(\lambda)$ and $\Sigma_0(\lambda, \lambda)$ when $\lambda \to -1$. The trend of $\Sigma_1(\lambda)$ is easy to derive. In fact, from the expression $\Sigma_1(\lambda)$ in Theorem 2, we have $\Sigma_1(-1) = E\dot{m}(X; \theta_0)\dot{m}^T(X; \theta_0)$. Let $\lambda_j = \lambda_l = \lambda$ in the second expression of $\Sigma_0(\lambda_j, \lambda_l)$ in Theorem 2, we can rewrite $\Sigma_0(\lambda, \lambda)$ as

$$E \iint \left\{ \sigma_\varepsilon^2 + [m(X, \theta_0) - m(X+u, \theta(\lambda))][m(X, \theta_0) - m(X+v, \theta(\lambda))] \right\} \dot{m}(X+u; \theta(\lambda))$$
$$\dot{m}^T(X+v; \theta(\lambda)) \phi(v, 0, (\lambda+1)\Sigma_u) \phi\left(u, \frac{v}{\lambda+1}, \frac{2\lambda+\lambda^2}{\lambda+1}\Sigma_u\right) du dv.$$

However, $\Sigma_0(\lambda, \lambda)$ does not have an explicit limit when $\lambda \to -1$, unless some strong conditions are imposed on the regression function.

Following are two examples of $\Sigma_0(\lambda, \lambda)$ to further illustrate the above findings.

Example 1. Suppose $m(x, \theta) = x'\theta$, then we have $\dot{m}(x, \theta) = x$. From the second expression of $\Sigma_0(\lambda, \lambda)$, we have

$$\begin{aligned}
\Sigma_0(\lambda, \lambda) = & \sigma_\varepsilon^2 E \iint (X+u)(X+v)^T \phi(v, 0, (\lambda+1)\Sigma_u) \phi\left(u, \frac{v}{\lambda+1}, \frac{\lambda(\lambda+2)}{\lambda+1}\Sigma_u\right) du dv \\
& + E \iint [X^T\theta_0 - (X+u)^T\theta(\lambda)][X^T\theta_0 - (X+v)^T\theta(\lambda)] \cdot \\
& \qquad (X+u)(X+v)^T \phi(v, 0, (\lambda+1)\Sigma_u) \phi\left(u, \frac{v}{\lambda+1}, \frac{\lambda(\lambda+2)}{\lambda+1}\Sigma_u\right) du dv \\
= & \sigma_\varepsilon^2 (EXX^T + \Sigma_u) + E(\theta_0 - \theta(\lambda))^T XX^T (\theta_0 - \theta(\lambda)) XX^T \\
& + (\theta_0 - \theta(\lambda))^T E(XX^T)(\theta_0 - \theta(\lambda))\Sigma_u + \theta(\lambda)^T \Sigma_u \theta(\lambda) E(XX^T) \\
& - E(\theta_0 - \theta(\lambda))^T XX\theta(\lambda)^T \Sigma_u - (\lambda+1) E(\theta_0 - \theta(\lambda))^T X \Sigma_u \theta(\lambda) X^T \\
& + \int v^T \theta(\lambda) \left[vv^T + \lambda(\lambda+2)\Sigma_u \right] \theta(\lambda) v^T \phi(v, 0, \Sigma_u) dv.
\end{aligned}$$

Extrapolating λ to -1, we see that $\Sigma_0(\lambda, \lambda)$ converges to

$$\sigma_\varepsilon^2 (EXX^T + \Sigma_u) + \theta_0^T \Sigma_u \theta_0 EXX^T + \int v^T \theta_0 v v^T \theta_0 v^T \phi(v, 0, \Sigma_u) dv - \Sigma_u \theta_0 \theta_0^T \Sigma_u. \qquad (2.3.3)$$

(2.3.3) is exactly same as the asymptotic covariance matrix of the bias corrected estimator $\hat{\beta}(\lambda)$ defined in (2.2.3) after taking $\lambda = -1$. Please refer to the Theorem 3.1 in Liang et al. (1999) for a proof of the asymptotic normality of the bias corrected estimator, as well as its asymptotic covariance matrix (2.3.3).

Example 2: Consider the exponential regression function $Y = \exp(\theta X) + \varepsilon$ with a univariate predictor X. Note that $\dot{m}(x;\theta) = x\exp(\theta x)$. It is easy to see that $\Sigma_1(\lambda) \to EX^2\exp(2X\theta_0) = \tau_1$ as $\lambda \to -1$. From the second expression for $\Sigma_0(\lambda,\lambda)$ in Theorem 2, we can rewrite it as the sum of the following two terms:

$$
\begin{aligned}
S_1(\lambda) = \sigma_\varepsilon^2 E \iint & (X+u)(X+v)\exp\left[(X+u)\theta(\lambda)\right]\exp\left[\theta(\lambda)(X+v)\right]\cdot \\
& \phi(v,0,(\lambda+1)\sigma_u^2)\phi\left(u,\frac{v}{\lambda+1},\frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\right)dudv, \\
S_2(\lambda) = E\iint & \{\exp(\theta_0 X) - \exp\left[\theta(\lambda)(X+u)\right]\}\cdot\{\exp(\theta_0 X) - \exp\left[\theta(\lambda)(X+v)\right]\}\cdot \\
& (X+u)(X+v)\exp\left[(X+u)\theta(\lambda)\right]\exp\left[\theta(\lambda)(X+v)\right]\cdot \\
& \phi(v,0,(\lambda+1)\sigma_u^2)\phi\left(u,\frac{v}{\lambda+1},\frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\right)dudv.
\end{aligned}
$$

A tedious computation shows that

$$
S_1(\lambda) = \sigma_\varepsilon^2 E\left\{(\sigma_u^2 + [X + (\lambda+2)\theta(\lambda)\sigma_u^2]^2)\exp[2X\theta(\lambda) + (\lambda+2)\theta^2(\lambda)\sigma_u^2]\right\},
$$

and

$$
\begin{aligned}
S_2(\lambda) = {} & EX^2\exp[2X(\theta_0+\theta(\lambda)) + (\lambda+2)\sigma_u^2\theta^2(\lambda)] \\
& -2EX^2\exp[X(\theta_0+3\theta(\lambda)) + \frac{1}{2}(5\lambda+9)\sigma_u^2\theta^2(\lambda)] \\
& +EX^2\exp[4X\theta(\lambda) + 4(\lambda+2)\sigma_u^2\theta^2(\lambda)] \\
& +2(\lambda+2)\sigma_u^2\theta(\lambda)EX\exp[2X(\theta_0+\theta(\lambda)) + (\lambda+2)\sigma_u^2\theta^2(\lambda)] \\
& -6(\lambda+2)\sigma_u^2\theta(\lambda)EX\exp[X(\theta_0+3\theta(\lambda)) + \frac{1}{2}(5\lambda+9)\sigma_u^2\theta^2(\lambda)] \\
& +4(\lambda+2)\sigma_u^2\theta(\lambda)EX\exp[4X\theta(\lambda) + 4(\lambda+2)\sigma_u^2\theta^2(\lambda)] \\
& +[\sigma_u^2 + (\lambda+2)^2\sigma_u^4\theta^2(\lambda)]E\exp[2X(\theta_0+\theta(\lambda)) + (\lambda+2)\sigma_u^2\theta^2(\lambda)] \\
& -2[\sigma_u^2 + (2\lambda+3)(\lambda+3)\sigma_u^4\theta^2(\lambda)]E\exp[X(\theta_0+3\theta(\lambda)) + \frac{1}{2}(5\lambda+9)\sigma_u^2\theta^2(\lambda)]
\end{aligned}
$$

$$+[\sigma_u^2 + 4(\lambda+2)^2\sigma_u^4\theta^2(\lambda)]E\exp[4X\theta(\lambda) + 4(\lambda+2)\sigma_u^2\theta^2(\lambda)].$$

Extrapolating λ to -1, we see that $\Sigma_0(\lambda,\lambda)$ tends to

$$\begin{aligned}
\tau_0^2 = \sigma_\varepsilon^2 E &\left\{(\sigma_u^2 + [X+\theta_0\sigma_u^2]^2)\exp[2X\theta_0 + \theta_0^2\sigma_u^2]\right\} \\
&+E\left\{(\sigma_u^2 + [X+\theta_0\sigma_u^2]^2)\exp[4X\theta_0 + \theta_0^2\sigma_u^2]\right\} \\
&+E\left\{[\sigma_u^2 + (X+2\theta_0\sigma_u^2)^2]\exp(4X\theta_0 + 4\theta_0^2\sigma_u^2)\right\} \\
&-2E\left\{(\sigma_u^2 + X^2 + 3X\theta_0\sigma_u^2 + 2\theta_0^2\sigma_u^4)\exp(4X\theta_0 + 2\sigma_u^2\theta_0^2)\right\}.
\end{aligned}$$

In fact, we can verify that the asymptotic variance of the estimator defined by the solution of (2.2.4) is exactly τ_0^2/τ_1^2.

It is noted that the exact extrapolation function $\theta(\lambda)$ is implicitly defined by the equation $\dot{L}(\theta;\lambda) = 0$, where $L(\theta;\lambda)$ is defined by (2.3.2). In some special cases, such as the linear, the exponential and the Poisson regressions discussed in Examples 1, 2 and 3, the exact extrapolation function can be obtained by solving the above equation. However, the solution generally has no closed-form.

For simplicity, we assume that θ is one-dimensional. By a Taylor expansion of $m(x + \sqrt{\lambda+1}\sigma_u u, \theta)$ and $\dot{m}(x + \sqrt{\lambda+1}\sigma_u u, \theta)$ at $\lambda = -1$ and θ at $\theta = \theta_0$, and plugging these two Taylor expansions in the expression of $\dot{L}(\theta;\lambda)$, and after some algebra, the solution of the equation $\dot{L}(\theta;\lambda) = 0$ has the form of

$$\theta(\lambda) = \theta_0 + \frac{a_0(\lambda+1)}{a_1 + a_2(\lambda+1) + a_3(\lambda+1)^2 + \cdots},$$

where $a_0, a_1, a_2, a_3, \ldots$ are some model-dependent constants. The above exact extrapolation function can be simplified to obtain several approximate simple functions to implement the extrapolation. For example, by truncating the denominator to the first order of $\lambda+1$, and after some equivalent transformation, we obtain the commonly used nonlinear extrapolation function $a + b/(c+\lambda)$; or if we apply another Taylor expansion for the denominator at

$\lambda = -1$ up to first order, second order etc., then we can also obtain the linear, quadratic extrapolation function and so on.

Almost all literature involving SIMEX assumes that the true extrapolation function has a known parametric form when discussing the asymptotic distributions of the SIMEX estimators. However, the true extrapolation function is generally unknown except for some special cases, and this discouraging observation really nullifies all the relevant theoretical developments based on known extrapolation functions. Unfortunately this is also true for our proposed extrapolation method.

If we are fortunate to have a closed-form extrapolation function with a parametric form $G(\lambda, \Gamma)$, which is twice continuously differentiable with respect to the parameter $\Gamma \in \mathbb{R}^d$ for some positive integer d, then assuming that the true value of the parameter is Γ_0, that is $\theta_0 = G(-1, \Gamma_0)$, we can estimate Γ_0 by minimizing the least squares criterion

$$\|\hat{\theta}_n(\Lambda) - G(\Lambda, \Gamma)\|_2^2 = \sum_{j=1}^{K} \|G(\lambda_j, \Gamma) - \hat{\theta}_n(\lambda_j)\|_2^2, \tag{2.3.4}$$

where $G(\Lambda, \Gamma) = [G^T(\lambda_1, \Gamma), G^T(\lambda_2, \Gamma), \ldots, G^T(\lambda_K, \Gamma)]^T_{qK\times 1}$, or solving the equation

$$\dot{G}^T(\Lambda, \Gamma)(\hat{\theta}_n(\Lambda) - G(\Lambda, \Gamma)) = 0,$$

where

$$\dot{G}(\Lambda, \Gamma) = [\dot{G}^T(\lambda_1, \Gamma), \dot{G}^T(\lambda_2, \Gamma), \ldots, \dot{G}^T(\lambda_K, \Gamma)]^T_{qK\times d},$$

and for $j = 1, 2, \ldots, K$,

$$\dot{G}(\lambda_j, \Gamma) = \left(\partial G(\lambda_j, \Gamma)/\partial \gamma_l\right)_{q\times d}.$$

Denote the minimizer of (2.3.4) as $\hat{\Gamma}$, then the resulting extrapolation estimator of θ_0 will be $\hat{\theta}_n = G(-1, \hat{\Gamma})$. Denote $H(\Lambda) = \dot{G}^T(\Lambda, \Gamma_0)\dot{G}(\Lambda, \Gamma_0)$ and

$$\Pi(\Lambda) = H^{-1}(\Lambda)\dot{G}^T(\Lambda, \Gamma_0)\Omega_1^{-1}(\Lambda)\Omega_0(\Lambda)\Omega_1^{-1}(\Lambda)\dot{G}(\Lambda, \Gamma_0)H^{-1}(\Lambda),$$

We have the following ideal result.

Theorem 3. *Assuming that conditions (C1)-C(6) hold, and the true extrapolation function has a known parametric form $G(\lambda, \Gamma)$ with nonsingular $H(\Lambda)$, then*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \Longrightarrow N(0, \dot{G}(-1, \Gamma_0)\Pi(\Lambda)\dot{G}^T(-1, \Gamma_0)).$$

2.4 Numerical Studies

In this section, we conduct some simulation studies for two parametric models to evaluate the finite sample performance of the proposed estimation procedure. For convenience, we label the proposed estimation procedure as EX, emphasizing the lack of the simulation step. In the first simulation study, we generated the data from a regression model with exponential regression function as described in Example 1. In the second simulation study, the data are generated from a bivariate quantile regression model as described in Example 6. Note that we can directly plug $\lambda = -1$ in the target function from the exponential regression model, but we cannot do the same in the quantile regression model, where an approximation is needed by using extrapolation functions. We also analyze a dataset from the National Health and Nutrition Examination Survey (NHANES) to illustrate the application of the proposed estimation procedure.

2.4.1 Simulation Studies

Exponential Regression. Consider a univariate exponential regression model first. Suppose X is a one-dimensional standard normal random variable. From Section 2.2, we know the true value θ_0 is a solution of the equation $EX\exp((\theta + \theta_0)X) - EX\exp(2\theta X) = 0$. Calculation shows that the equation can be rewritten as

$$(\theta + \theta_0)\exp((\theta + \theta_0)^2/2) - 2\theta\exp(2\theta^2) = 0.$$

It is easy to see that θ_0 is a solution of the above equation. Moreover, the solution of the above equation is also unique. Therefore, when the sample size is sufficiently large, the minimizer of $L_n(\theta)$ is unique. We generate Z from $Z = X + U$ with $U \sim N(0, \sigma_u^2)$. To see the effect of the measurement error variance on the estimation procedure, we choose $\sigma^2 = 0.1, 0.25, 0.5$. The true parameter values are chosen to be $\theta = 1$, and the sample sizes are chosen to be $n = 200, 300, 500$ and 800. For each setup, we repeat the simulation 500 times, and the mean squared errors (MSE) are calculated and used to evaluate the finite sample performance of the EX. For comparison, the traditional SIMEX with $B = 50$ is also applied to estimate the unknown parameter θ. Along with the MSE values, we also report the computation time to obtain these MSE values.

Table 2.1: MSEs and Computation Time (minutes): univariate exponential regression

	n=200			n=300			n=500			n=800		
σ_u^2	0.5	0.25	0.1	0.5	0.25	0.1	0.5	0.25	0.1	0.5	0.25	0.1
EX												
MSE	0.029	0.017	0.005	0.019	0.012	0.003	0.010	0.007	0.003	0.008	0.005	0.002
Time	0.344	0.350	0.351	0.472	0.483	0.487	0.751	0.808	0.760	1.132	1.138	1.154
SIMEX												
MSE	0.011	0.010	0.004	0.008	0.006	0.003	0.006	0.005	0.003	0.005	0.003	0.002
Time	4.818	4.681	4.632	5.068	5.053	4.997	5.920	5.869	5.799	7.223	7.154	7.020

Table 2.1 shows that when the sample size gets bigger, the performance of the EX estimator becomes better, as evidenced by the decreasing MSEs. For a fixed sample size, it is seen that the smaller the measurement error variance, the smaller the MSEs, as expected. For comparison, we also conduct a simulation using the SIMEX procedure. The computation time (in minutes) for both estimation procedures is also reported in Table 2.1. From Table 2.1, we can see that the EX and the SIMEX performs similarly, but the latter is more computationally intensive than the former.

We also conducted a simulation study for an exponential regression with two predictors $X = (X_1, X_2)$, that is, $m(x; \theta) = \exp(x^T\theta)$, $Z_1 = X_1 + U_1, Z_2 = X_2 + U_2$, where $\theta = (\beta_1, \beta_2)$. We choose ε to be standard normal, (X_1, X_2) has a bivariate standard normal distribution, the measurement error (U_1, U_2) is generated from a bivariate normal distribution with mean vector of 0's, equal variances σ^2 and covariance $0.5\sigma^2$. To see the effect of the measurement variance on the estimation procedure, we choose $\sigma^2 = 0.25, 0.2, 0.1$. The true parameter values are chosen to be $\beta_1 = 0.5$, $\beta_2 = 1$, and the sample sizes are chosen to be $n = 200, 300, 500$ and 800. Similar to the one dimensional case, for each setup, we repeat the simulation 500 times, and the MSEs are calculated and used to evaluate the finite sample performance of the proposed EX procedure and the traditional SIMEX procedure.

Table 2.2: MSEs and Computation Time (minutes): bivariate exponential regression

	n=200						n=300					
σ_u^2	0.25		0.2		0.1		0.25		0.2		0.1	
EX												
MSE	0.116	0.210	0.067	0.078	0.017	0.014	0.088	0.107	0.058	0.040	0.015	0.009
Time	7.124	7.124	7.124	7.124	7.168	7.168	9.410	9.410	9.460	9.460	9.407	9.407
SIMEX												
MSE	0.033	0.036	0.024	0.029	0.013	0.012	0.029	0.033	0.024	0.027	0.012	0.010
Time	17.827	17.827	17.779	17.779	15.834	15.834	19.620	19.620	19.017	19.017	17.631	17.631

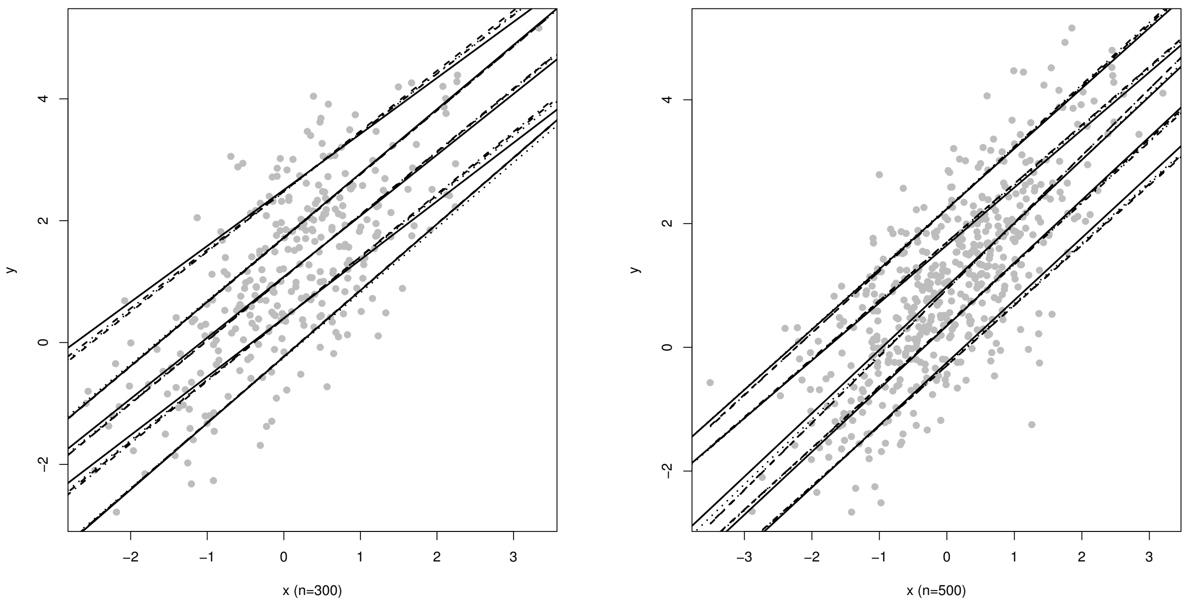
Table 2.3: MSEs and Computation Time (minutes): bivariate exponential regression

	n=500						n=800					
σ_u^2	0.25		0.2		0.1		0.25		0.2		0.1	
EX												
MSE	0.060	0.079	0.037	0.026	0.011	0.006	0.045	0.033	0.025	0.015	0.006	0.004
Time	13.95	13.95	13.992	13.992	14.317	14.317	21.009	21.009	21.081	21.081	20.993	20.993
SIMEX												
MSE	0.028	0.030	0.019	0.024	0.010	0.009	0.021	0.029	0.017	0.022	0.009	0.008
Time	24.418	24.418	23.105	23.105	21.531	21.531	31.46	31.46	30.552	30.552	27.249	27.249

From Table 2.2 and 2.3, we can see the EX has slightly larger MSE values than the SIMEX procedure, but its computational time is much less than the SIMEX procedure.

Quantile Regression. In this simulation study, we consider a quantile regression model with univariate predictor, as discussed in Example 6 in Section 2. The simulated data are generated from the linear regression model $Y = \beta_0 + \beta_1 X + \varepsilon$, where X has a standard normal distribution. For ε, we consider two distributions – the standard normal distribution and $(\chi_2^2 - 1.3863)/2$, where χ_2^2 denotes a χ^2-distribution with degrees of freedom 2. Note that 1.3863 and 2 are the 50-th percentile and the standard deviation of the χ^2-distribution, respectively, so ε has a median of 0 and variance 1. In the simulation study, the sample size is chosen to be $n = 300$, and $n = 500$, and two measurement error variances, 0.1^2 and 0.5^2, are used to evaluate the effect of the measurement error on the quantile regression line. The quadratic function is used for extrapolation.

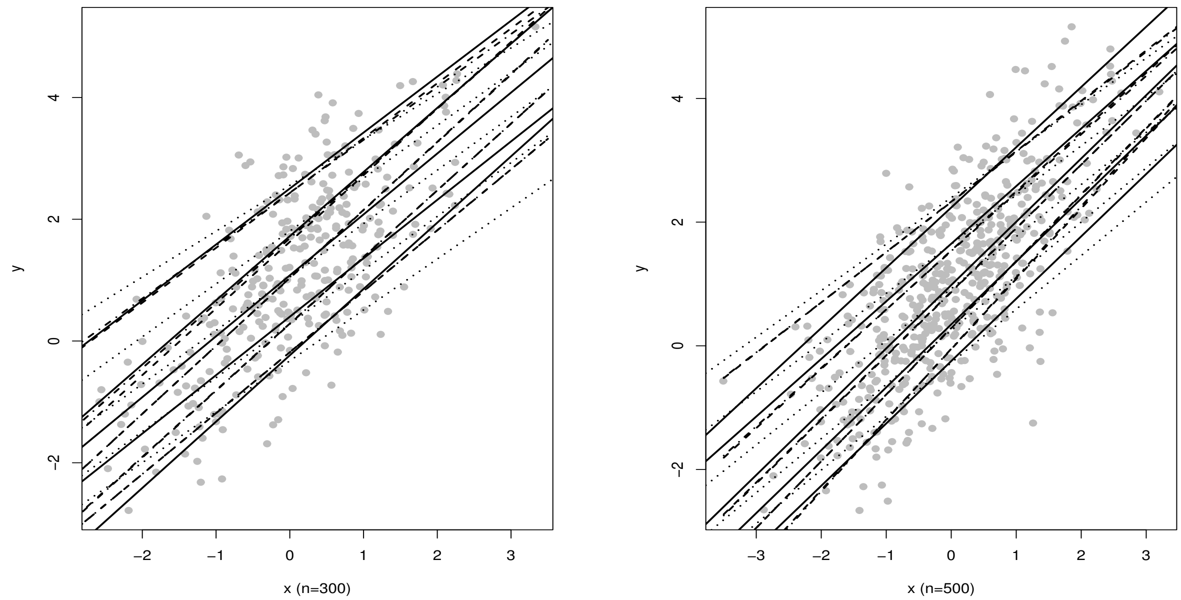
Figure 2.1: Quantile Regression (Normal, $\sigma_u = 0.1$)



The dashed line in the left plot on Figure 2.1 shows the fitted quantile regression lines using the EX method when $n = 300$, $\sigma_u = 0.1$, and ε is normally distributed. From the top to the bottom, the dashed lines are corresponding to the fitted $90, 75, 50, 25$ and 10-th quantiles lines. For comparison purposes, we also plot the quantile regression lines fitted using the SIMEX method (dash-dotted lines), the naive method (the dotted lines), and as a benchmark, the quantile regression lines fitted from the data on (Y, X) (solid lines) are also plotted. These fitted quantile lines are well-matched to the benchmark lines. Except for some slight deviations, all methods perform equally well, including the naive method. The right plot on Figure 2.1 is for $n = 500$, similar patterns are observed. Figure 2.2 shows the simulated quantile regression lines for $\sigma_u = 0.5$, and the superiority of the EX method and the traditional SIMEX over the naive method is observed. We can also see that the performance of the EX method and the traditional SIMEX is comparable.

The results from the simulation studies when ε follows $(\chi_2^2 - 1.3863)/2$ are shown in Fig-

Figure 2.2: Quantile Regression (Normal, $\sigma_u = 0.5$)



ures 2.3 and 2.4. Note that in all scenarios except for $\sigma_u = 0.1$, it seems like the EX method and the traditional SIMEX are not very satisfying when estimating the 90-th quantile. This might be partially due to the fact that a right-skewed error distribution tends to generate more large outliers. When σ_u is large, both the EX method and the traditional SIMEX outperform the naive method in general.

2.4.2 Real Data Application

To investigate whether the serum 25-hydroxyvitamin D (25(OH)D) is influenced by the long term vitamin D average intake or not, Curley (2017) analyzed a data set from the National Health and Nutrition Examination Survey (NHANES), and used a nonlinear function to model the regression mean of 25(OH)D on the long term vitamin D average intake. In this section, we apply the proposed estimation procedure on a subset of the 2009-2010 NHANES study. The selected data set contains dietary records of 806 Mexican American females. The long term vitamin D average intake X is not measured directly, instead, two independent daily observations of vitamin D intake are collected. Let Z_{ji} be the vitamin D intake from the

Figure 2.3: Quantile Regression (χ^2, $\sigma_u = 0.1$)

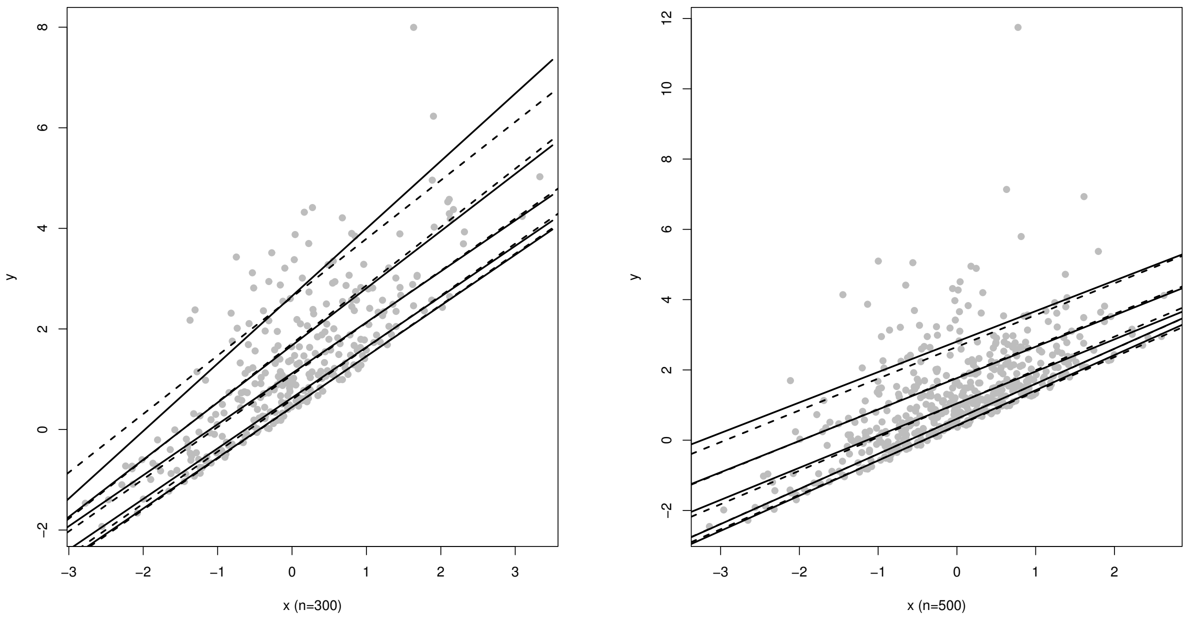
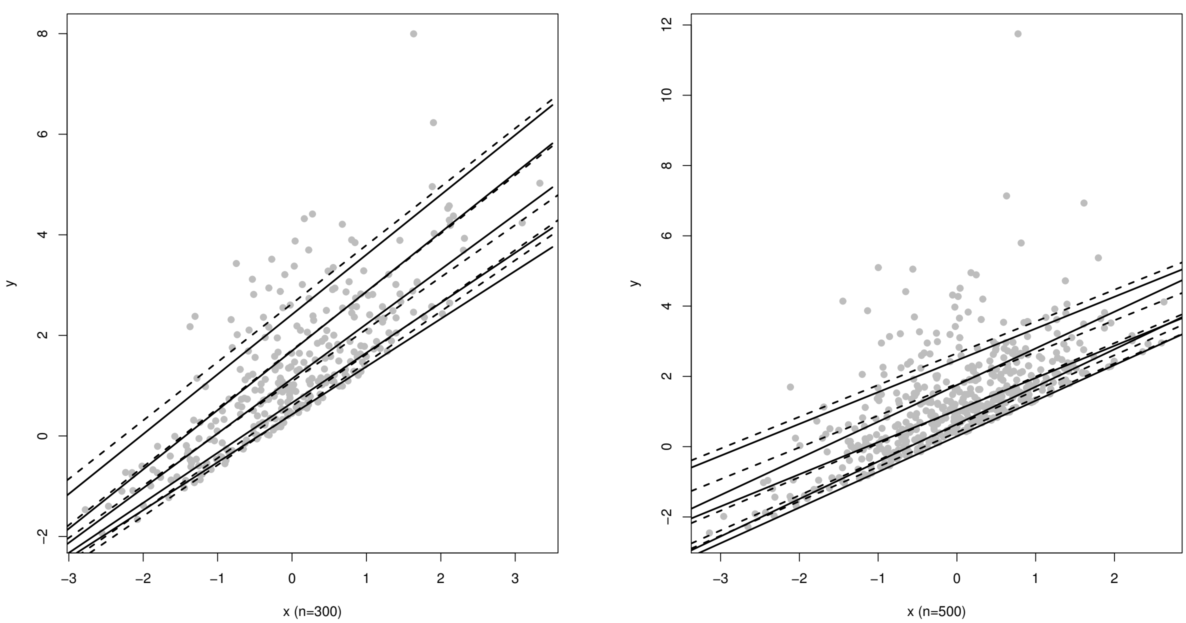


Figure 2.4: Quantile Regression (χ^2, $\sigma_u = 0.5$)



i-th subject at the j-th time, and we assume that the additive structures $Z_{ji} = X_i + U_{ji}$ hold for all $i = 1, 2, \ldots, 806$, $j = 1, 2$. We use $Z_i = (Z_{1i} + Z_{2i})/2$ to represent the observed vitamin intake, and by assuming that U_{1i} and U_{2i} are independently and identically distributed, we can estimate the standard deviation of the measurement error U by the sample standard deviation of the differences $(Z_{1i} - Z_{2i})/2$, $i = 1, 2, \ldots, n$. Similar to Curley (2017), a square root transformation of the 25(OH)D is used to achieve a more symmetric structure.

We adopt the S-shaped function $\beta_0 + \beta_1/(1 + \exp\{\beta_2(X - \beta_3)\})$ as the regression function of Y against X, which is also used in Curley (2017). 11 equally spaced values are chosen from $[0, 1]$ as the λ values. By checking the estimates of the parameters at various λ-values, as the circles shown in Figure 2.5, we can see that the quadratic extrapolation function seems more proper. The EX estimates using the quadratic extrapolation are $\beta_0 = 7.462$, $\beta_1 = -1.598$, $\beta_2 = 0.742$, $\beta_3 = -2.515$ and the solid line in Figure 2.6 represents the fitted regression function. For comparison purpose, we also fit the model using linear extrapolation function, as the dashed line shown in Figure 2.6. Clearly, the fitted regression function based on the quadratic extrapolation function seems to fit the data structure better than the one based on the linear extrapolation function. We also applied the traditional SIMEX method with $B = 100$ to fit the data, also using the quadratic extrapolation function, the result is shown as the dotted line in Figure 2.6, and one can see that it behaves similarly to the EX fitted lines using the quadratic extrapolation. Finally, we also fit the data using the naive method, and the estimates are $\beta_0 = 8.077$, $\beta_1 = -1.626$, $\beta_2 = 0.322$, $\beta_3 = 3.864$. Note that the naive estimate for β_3 has a different sign from the EX estimate.

Figure 2.5: Extrapolation plots using quadratic function

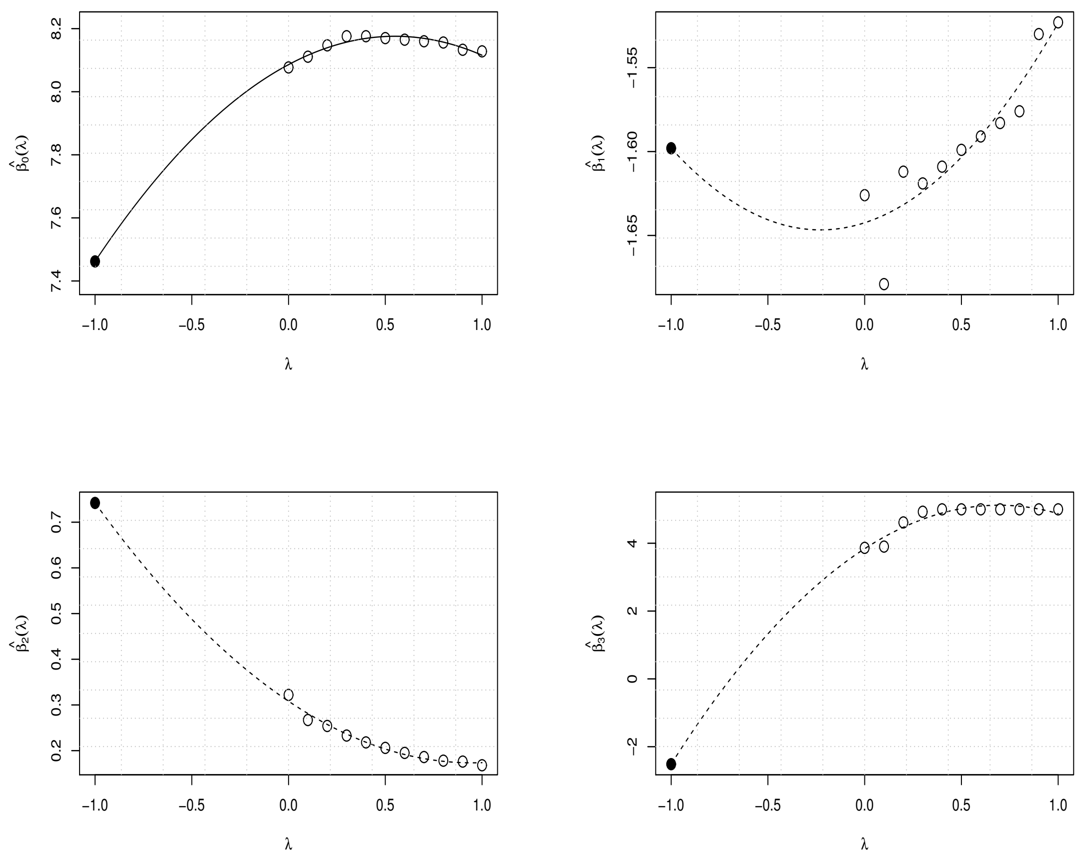
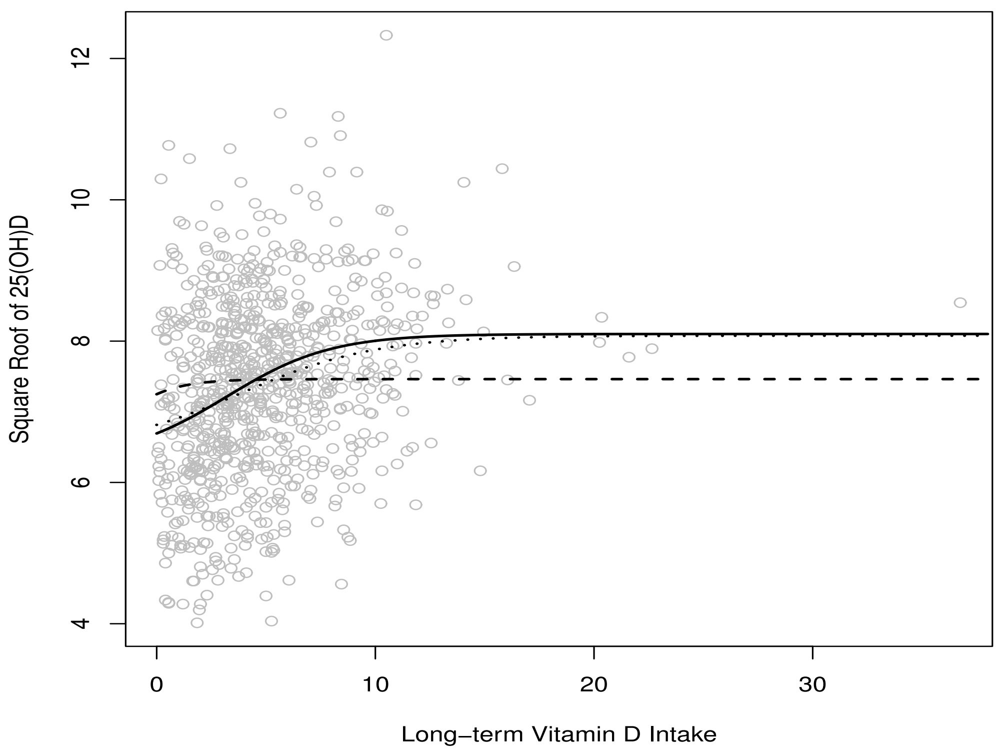


Figure 2.6: Nonlinear regression for HNANES data. The solid line represents the fit from the proposed method using quadratic extrapolation, the dashed line represents the fit from the proposed method using linear extrapolation and the dotted line denotes the fit from the SIMEX procedure.



2.5 Discussion

The extrapolation estimation procedure proposed in this paper has a potential extension to other more complicated regression models when some predictors are contaminated with normal measurement errors, such as the partially linear regression models, the varying coefficients regression models or other semi-parametric models, where the error-prone variables appear as a linear component.

The extension to the partially linear regression model $Y = X^T\beta + g(T) + \varepsilon$ is straightforward, where X is a p-dimensional latent predictor, g is an unknown univariate function satisfying some smoothness conditions. Suppose $Z = X + U$, $U \sim N(0, \Sigma_u)$, and T can be observed directly. Then, using the least squares procedure, and following the protocol of the extrapolation algorithm described in Section 2.3, we can estimate β and g by minimizing the conditional expectation $E\left[\sum_{i=1}^n [Y_i - Z_i^T(\lambda)\beta - g(T_i)]^2 | \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{T}\right]$ or $\sum_{i=1}^n [Y_i - Z_i^T\beta - g(T_i)]^2 + \lambda\beta^T\Sigma_u\beta$, where $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$, $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^T$, and $\boldsymbol{T} = (T_1, \ldots, T_n)^T$. Note that $\lambda = -1$ can be directly plugged in. Applying the profile least squares procedure, that is replacing $g(T_i)$ by its pseudo-kernel estimate $\hat{g}(t; \beta) = \sum_{j=1}^n L_h(t - T_j)(Y_j - Z_j^T\beta) / \sum_{j=1}^n L_h(t - T_j)$, where $L_h(\cdot) = L(\cdot/h)/h$ for a kernel function L and a bandwidth h, we can estimate β by the minimizer of $\sum_{i=1}^n [Y_i - Z_i^T\beta - \hat{g}(T_i)]^2 - \beta^T\Sigma_u\beta$ and eventually estimate $g(t)$ with $\hat{g}(t; \hat{\beta})$. This is the same estimation procedure used in Liang et al. (1999) when they discuss the variable selection for the partially linear models with measurement errors.

The varying coefficient regression model assumes that a scalar response variable Y depends on the explanatory vectors X and W via the relationship $Y = X^T g(W^T\beta) + \varepsilon$, where X is a latent vector following the measurement error structure $Z = X + U$ with $U \sim N(0, \Sigma_u)$, and W is observable. The unknown vector function g and the parameter β are the quantities of interest. Pretending g is known, then we can estimate β by minimizing the conditional expectation $E\left[\sum_{i=1}^n [Y_i - Z_i^T(\lambda) g(W_i^T\beta)]^2 | \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{W}\right]$ or $\sum_{i=1}^n [Y_i - Z_i^T g(W_i^T\beta)]^2 + \lambda g^T(W_i^T\beta)\Sigma_u g(W_i^T\beta)$. Similar to the partially linear regression case, one can apply the two-step profile least squares estimation procedure to estimate g and β after replacing λ with -1.

Another important question is the robustness of proposed extrapolation procedure against the misspecification of the normality assumption on the measurement error. Many researchers have claimed that the classcial SIMEX procedure is robust. However, Koul and Song (2014) proved theoretically, along with some examples, that this statement is not true. In fact, we can see that the procedure proposed in this paper is not robust either. To see this, let us revisit the Poisson regression model discussed in Section 2.2 with a univariate predictor X. Suppose the measurement error U is normally distributed and the true parameter value is θ_0, then by adding extra normal error, we have $E[YZ\theta - \exp(Z\theta)\exp(\lambda\theta^2\sigma_u^2/2)]$ equals

$$E[YX\theta - \exp(X\theta)\exp((\lambda+1)\theta^2\sigma_u^2/2)] = E[X\theta\exp(X\theta_0) - \exp(X\theta)\exp((\lambda+1)\theta^2\sigma_u^2/2)].$$

Extrapolating $\lambda \to -1$, one can see that θ_0 is a maximizer of $E[X\theta\exp(X\theta_0) - \exp(X\theta)]$. But if U has a Laplace distribution with mean 0 and variance σ_u^2, and we still proceed by adding extra normal errors, then, note that for such a Laplace random variable U, we have $E\exp(U\theta) = (1 - \sigma_u^2\theta^2)^{-1}$, so

$$E[YZ\theta - \exp(Z\theta)\exp(\lambda\theta^2\sigma_u^2/2)] = E[YX\theta - \exp(X\theta)(1 - \sigma_u^2\theta^2/2)^{-1}\exp(\lambda\theta^2\sigma_u^2/2)].$$

If we also assume that $X \sim N(0, \sigma_u^2)$, then extrapolating $\lambda \to -1$, the above expectation becomes $\theta\theta_0\sigma_u^2\exp(\theta_0^2\sigma_u^2/2) - (1 - \sigma_u^2\theta^2/2)^{-1}$, and its maximum does not achieve at $\theta = \theta_0$ in general.

To conclude this section, we would like to point out that the proposed method provides an alternative way to look for the corrected score function defined in Nakamura (1990), if the conditional expectation (2.3.1) can be directly extrapolated to $\lambda = -1$. Also, we must be aware of the limitation of the proposed EX procedure. If the target conditional expectations, such as the ones defined in, have a more complicated expression than the target functions used in the traditional SIMEX, the efficiency of EX procedure gained from skipping the simulation step in the traditional SIMEX might be compromised by a time-consuming step of optimization. Therefore, the proposed EX procedure should be viewed as a complement to the SIMEX but not a replacement.

Chapter 3

Extrapolation Estimation for Nonparametric Regression with Measurement Error

3.1 Introduction

Due to its conceptual simplicity and the capability to harness the modern computational power, the simulation extrapolation estimation (SIMEX) procedure has been attracting significant attention from practical data analysts as well as theoretical researchers. The simplicity of the SIMEX lies in the fact that it allows us to directly use any standard estimates based on the known data as the building block, and its simulation nature makes the estimation process computer-dependent only. To be specific, suppose we want to estimate a parameter θ, possibly multidimensional, in a statistical population X of dimension p, where $p \geq 1$. In certain situations where we cannot collect observations directly from X, what we observe is a surrogate value Z of X. In measurement error literature, a classical assumption on the relationship between X and Z is $Z = X + U$, where U is called the measurement error, which is often assumed to be independent of X, and has a normal distribution with mean 0 and known covariance matrix Σ_u. If there is an estimator $T(\boldsymbol{X})$ of θ when a sample

$\boldsymbol{X} = \{X_1, \ldots, X_n\}$ of size n from X is available, then when only Z can be observed, the classical SIMEX procedure estimates θ, using sample $\boldsymbol{Z} = \{Z_1, \ldots, Z_n\}$ from Z, by going through the following three steps. First, we generate n i.i.d. random vectors V_i's from $N(0, \Sigma_u)$, select a nonnegative number λ, calculate $\tilde{Z}_i(\lambda) = Z_i + \sqrt{\lambda} V_i$ for $i = 1, 2, \ldots, n$, and compute $T(\tilde{\boldsymbol{Z}}(\lambda))$ based on $\tilde{\boldsymbol{Z}}(\lambda) = \{\tilde{Z}_1(\lambda), \ldots, \tilde{Z}_n(\lambda)\}$. Second, we calculate the conditional expectation of $T(\tilde{\boldsymbol{Z}}(\lambda))$ given $\boldsymbol{Z}$. If the conditional expectation has a closed form, then it will be the estimate of θ, otherwise, we repeat the previous step B times to obtain B values of $T_b(\tilde{\boldsymbol{Z}}(\lambda))$, $b = 1, 2, \ldots, B$, and the average $\bar{T}(\lambda)$ of these B values of $T_b(\tilde{\boldsymbol{Z}}(\lambda))$'s is computed. Finally, we repeat the first step and second step for a sequence of nonnegative λ values, for example, $0 = \lambda_1 < \ldots < \lambda_K$ for some K. We denote these K averages as $\bar{T}(\lambda_1), \ldots, \bar{T}(\lambda_K)$. To conclude, the trend of $\bar{T}(\lambda)$ with respect to λ will be formulated as a function of λ, and the extrapolated value of this function at $\lambda = -1$ is the desired SIMEX estimate of θ. In real applications, K is suggested to be less than 20 and these K λ-values are chosen equally spaced from $[0, 2]$. The early development of the classical SIMEX estimation procedure can be found in Cook and Stefanski (1994), Stefanski and Cook (1995) and Carroll et al. (1996), with extensive applications in Mallick et al. (2002) for cox regression, Sevilimedu et al. (2019) for Log-logistic accelerated failure time models, Gould et al. (1999) for the catch-effort analysis, Hwang and Huang (2003), Stoklosa et al. (2016) for the capture-recapture models, Lin and Carroll (1999) for the analysis of the Framingham heart disease data using the logistic regression, Hardin et al. (2003) for generalized linear models, and Ponzi et al. (2019) for some applications in ecology and evolution, to name a few.

However, the discussion of the classical SIMEX estimation procedure in the nonparametric setup seems scant in the literature. Stefanski and Bay (1996) applied the simulation extrapolation procedure to estimate the cumulative distribution function of a finite population based on the Horvitz-Thompson estimator. Since the conditional expectation of the Horvitz-Thompson estimator with the true variable replaced by the pseudo-data given the observable surrogates has an explicit form, the simulation step can be bypassed. Also, the quadratic function of λ is shown to be a reasonable extrapolation function. Carroll et al. (1999) extended the classical SIMEX procedure to the nonparametric regression setup and it was implemented with the local linear estimator. In Carroll et al. (1999)'s work, the three steps in the classical SIMEX procedure are strictly followed. To estimate the unknown variance function in a general one-way analysis of variance model, Carroll and Wang (2008) proposed a permutation SIMEX estimation procedure to completely remove the bias after extrapolation. Wang et al. (2010) generalized Stefanski and Bay (1996)'s method to estimate the smooth distribution function in the presence of heteroscedastic normal measurement errors. Aiming at improving the SIMEX local linear estimator in Carroll et al. (1999), Staudenmayer and Ruppert (2004) introduced a new local polynomial estimator with the SIMEX algorithm. The improvement over the existing estimation procedure is made possible by using a bandwidth selection procedure. Again, Staudenmayer and Ruppert (2004)'s method still strictly followed the three-steps in the classical SIMEX.

Compared to various applications in both the parametric and nonparametric statistical models, the SIMEX procedure developed in Stefanski and Bay (1996) and Wang et al. (2010) successfully dodged the simulation step, which is the most time-consuming part in the classical SIMEX algorithm. The very reason why their methods work is that the averaged naive estimator from the pseudo-data, conditioning on the observed surrogates, has an explicit limit ready for extrapolation, as the number of pseudo-data sets tends to infinity. Clearly, the strategy used in both references cannot be directly extended to other scenarios where such limits do not have user-friendly forms. In this paper, we will propose a new method, which in spirit is a variant of the classical SIMEX procedure, for estimating the nonparametric regression. The new method can also successfully circumvent the simulation step, and the applicable extrapolation functions can also be found, although still being approximated, based on the true but not usable extrapolation functions derived from the theory.

3.2 Motivating Examples

In this section, we shall discuss two motivating examples which inspired our interest in searching for a more efficient bias reduction estimation procedure in the nonparametric setup. Our ambition is to keep the attractive feature of the extrapolation component in the classical SIMEX algorithm, while at the same time, significantly reducing the computational burden.

3.2.1 Simple linear regression model

Let Y and X be two univariate random variables, which obey a simple linear relationship $E(Y|X) = \alpha + \beta X$. Suppose we cannot observe X but we have data on $Z = X + U$ and $U \sim N(0, \sigma_u^2)$ with σ_u^2 being known. As discussed in Carroll et al. (1999), for any fixed $\lambda > 0$, after repeatedly adding the extra measurement errors, and computing the ordinary least squares slope, the averaged estimator consistently estimates $g(\lambda) = \beta\sigma_X^2/(\sigma_X^2 + (1+\lambda)\sigma_u^2)$. Obviously, extrapolating λ to -1, we have $g(-1) = \beta$. This clearly shows that SIMEX works very well for linear regression model. In fact, in the seminal paper Cook and Stefanski (1994), the SIMEX estimators of α and β can be derived without the simulation step. However, the derivation relies on a notion of NON-IID pseudo-errors. More details about the NON-IID pseudo-errors can be found in Cook and Stefanski (1994) and Section 5.3.4.1 in Carroll et al. (2006). Here we would like to point out that the SIMEX estimators of α, β can be obtained without using the NON-IID pseudo-errors.

Recall that the least squares (LS) estimator of α and β can be obtained by minimizing the LS criterion $\sum_{i=1}^n (Y_i - \alpha - \beta X_i)^2$. Since X_i are not available, following the SIMEX idea, we generate the pseudo-data $Z_i(\lambda) = Z_i + \sqrt{\lambda} V_i$, $i = 1, 2, \ldots, n$. However, instead of following the classical SIMEX road map to minimize the LS target function $\sum_{i=1}^n (Y_i - \alpha - \beta Z_i(\lambda))^2$, we minimize the conditional expectation $E\left[\sum_{i=1}^n (Y_i - \alpha - \beta Z_i(\lambda))^2 \Big| \boldsymbol{D}\right]$, where $\boldsymbol{D} = (\boldsymbol{Y}, \boldsymbol{Z})$, $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ and $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$. Since V_i's are i.i.d. from $N(0, \sigma_u^2)$ and independent of other random variables in the model, so this conditional expectation equals $\sum_{i=1}^n (Y_i - \alpha - \beta^T Z_i)^2 + n\lambda \boldsymbol{\beta}^T \Sigma_U \boldsymbol{\beta}$. The minimizer of the above expression is simply $\hat{\beta}(\lambda) = (S_{ZZ} + \lambda\Sigma_U)^{-1} S_{YZ}$ and $\hat{\alpha}(\lambda) = \bar{Y} - \hat{\beta}^T(\lambda)\bar{X}$ and by choosing $\lambda = -1$, we immediately have the commonly used bias-corrected estimators or the SIMEX estimators derived using NON-IID pseudo-errors. Note that here not only do we not need the simulation step, but also the extrapolation step is unnecessary.

3.2.2 Kernel density estimation

Suppose we want to estimate the density function $f_x(x)$ of X in the measurement error model $Z = X + U$. When observations can be made directly on X, the kernel density estimation procedure is often called on for this purpose. Starting with the classical kernel estimator, Wang et al. (2009) followed the classical SIMEX algorithm, constructed an average of the kernel estimator $\hat{f}_{B,n}(x) = B^{-1}\sum_{b=1}^{B}[n^{-1}\sum_{i=1}^{n} K_h(x - Z_i - \sqrt{\lambda}V_{i,b})]$ with B pseudo-data sets $\{Z_i + \sqrt{\lambda}V_{i,b}\}_{i=1}^{n}$, $b = 1, 2, \ldots, B$, where $K_h(\cdot) = h^{-1}K(\cdot/h)$. By the law of large numbers, $\hat{f}_{B,n}(x) \to n^{-1}\sum_{i=1}^{n}\int K_h(x - Z_i - \sqrt{\lambda}\sigma_u u)\phi(u)du = \tilde{f}_n(x)$ in probability. After some algebra, Wang et al. (2009) proposed to estimate $f_x(x)$ using $\hat{f}_n(x) = n^{-1}\sum_{i=1}^{n}(\sqrt{\lambda}\sigma_u)^{-1}\phi((x - Z_i)/\sqrt{\lambda}\sigma_u)$ which approximates the limit $\tilde{f}_n(x)$ for sufficiently large n. In fact, before initiating the simulation step, Cook and Stefanski (1994) suggested one should try to calculate the conditional expectation $E[f_{B,n}(x)|\boldsymbol{Z}]$ first. If this conditional expectation has a tractable form, then it will be chosen as the SIMEX estimator. Clearly, the conditional expectation is simply $\tilde{f}_n(x)$. It is interesting to note that if we deliberately choose the kernel function K to be standard norm, we can show that $\tilde{f}_n(x) = (n\sqrt{\lambda\sigma_u^2 + h^2})^{-1}\sum_{i=1}^{n}\phi((x - Z_i)/\sqrt{\lambda\sigma_u^2 + h^2})$ which can also be directly used for extrapolation. Because there is no approximation done here, $\tilde{f}_n(x)$ should potentially perform better than the estimator $\hat{f}_n(x)$ as proposed in Wang et al. (2009).

It is easy to see that the technique used in the kernel density estimation cannot be extended to the regression setup, since the commonly used kernel regression estimators, either the Nadaraya-Watson estimator, or the local linear estimator, often appear as a fraction of kernel components, which fails to provide a tractable conditional expectation for direct extrapolation. However, the observation of recovering the commonly used bias-corrected estimators or the SIMEX estimators derived using NON-IID pseudo-errors in the linear errors-in-variables regression indicates that we could have some interesting findings if we can apply the conditional expectation argument directly on the target functions, instead of computing the conditional expectation of the resulting naive estimator. In the next section, we will implement this idea via estimating the nonparametric regression function using a local linear smoothing procedure.

3.3 Extrapolation Estimation Procedure via Local Linear Smoother

For the sake of simplicity, we restrict ourselves to the univariate predictor cases. The proposed methodology can handle the multivariate predictor cases very well at the cost of introducing more complex notations. To be specific, suppose that the random pair (X, Y) obeys the following nonparametric regression model

$$Y = g(X) + \varepsilon, \quad Z = X + U \tag{3.3.1}$$

with the common assumption on ε, $E(\varepsilon|X) = 0$ and $0 < \tau^2(X) = E(\varepsilon^2|X) < \infty$. X and U are independent and U has a normal distribution $N(0, \sigma_u^2)$ with known σ_u^2. If (X, Y) are available, the local linear estimator for $g(x)$ at a fixed x-value in the domain of X is defined as

$$\hat{g}_n(x) = \frac{S_{2n}(x)T_{0n}(x) - S_{1n}(x)T_{1n}(x)}{S_{2n}(x)S_{0n}(x) - S_{1n}^2(x)},$$

where $S_{jn}(x) = n^{-1}\sum_{i=1}^{n}(X_i - x)^j K_h(X_i - x), T_{jn}(x) = n^{-1}\sum_{i=1}^{n}(X_i - x)^j Y_i K_h(X_i - x)$, and $j = 0, 1, 2$ for $S_{jn}(x)$, $j = 0, 1$ for $T_{jn}(x)$, $K_h(\cdot) = h^{-1}K(\cdot/h)$, and K is a kernel function, h is a sequence of positive numbers often called bandwidths. In the measurement error setup, a classical SIMEX estimator of g can be obtained through three steps: simulation, estimation and extrapolation. For the sake of completeness, the following algorithm provides a detailed guideline for implementing the three steps in estimating $g(x)$ from data on Y, Z.

-

SIMEX Algorithm of Local Linear Smoother

(1) Pre-select a sequence of positive numbers $\lambda = \lambda_1, \dots, \lambda_K$.

(2) For $\lambda = \lambda_1$, repeat the following steps B times. At the b-th repetition,

(i) Generate n i.i.d. random observations $V_{i,b}$'s from $N(0, \Sigma)$, and calculate $Z_{i,b}(\lambda) = Z_i + \sqrt{\lambda_1} V_i = X_i + U_i + \sqrt{\lambda_1} V_{i,b}$, $i = 1, 2, \dots, n$.

(ii) Compute

$$\hat{g}_{n,b}(x; \lambda_1) = \frac{S_{2nb}(x) T_{0nb}(x) - S_{1nb}(x) T_{1nb}(x)}{S_{2nb}(x) S_{0nb}(x) - S_{1nb}^2(x)},$$

where

$$S_{jnb}(x) = \frac{1}{n} \sum_{i=1}^{n} (Z_{i,b}(\lambda) - x)^j K_h(Z_{i,b}(\lambda) - x), \quad j = 0, 1, 2,$$
$$T_{lnb}(x) = \frac{1}{n} \sum_{i=1}^{n} (Z_{i,b}(\lambda) - x)^l Y_i K_h(Z_{i,b}(\lambda) - x), \quad l = 0, 1.$$

(3) Calculate $\hat{g}_{n,B}(x; \lambda_1) = B^{-1} \sum_{b=1}^{B} \hat{g}_{n,b}(x; \lambda_1)$.

(4) Repeat (2)-(3) for $\lambda = \lambda_2, \dots, \lambda_K$.

(5) Identify a parametric trend of the pairs $(\lambda_k, \hat{g}_{n,B}(x; \lambda_k))$, $k = 1, 2, \dots, K$ and denote the trend as a function $\Gamma(x; \lambda)$. The SIMEX estimator of g is defined as $\hat{g}_{\text{SIMEX}}(x) = \Gamma(x; -1)$.

As a rough guideline, the λ values are often selected as a sequence of equally spaced grid points from $[0, 2]$, K is a positive integer as small as 5 or as large as 20, and B is often chosen to be 100 or above. With such choices, one can see the classical SIMEX procedure for implementing the local linear smoother is computationally intensive.

To introduce our estimation procedure, we start with the local linear procedure. If X

can be observed, then based on a sample $(X_i, Y_i), i = 1, 2, \ldots, n$ from model (3.3.1), the local linear estimator of the regression function g, as well as its first order derivative at x, can be obtained by minimizing the following target function $L(\beta_0, \beta_1) = \sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1(X_i - x))^2 K_h(X_i - x)$ with respect to β_0 and β_1. In fact, the solution of β_0 is the local linear estimator of $g(x)$ and β_1 is the local linear estimator of $g'(x)$.

For a positive constant λ, we replace X_i with the pesudo-data $Z_i(\lambda) = Z_i + \sqrt{\lambda}V_i$ in the weighted least squares $L(\beta_0, \beta_1)$, and calculate its conditional expectation given $(Z_i, Y_i), i = 1, 2, \ldots, n$. A straightforward calculation shows that the minimizer of

$$\sum_{i=1}^{n} E\left([Y_i - \beta_0 - \beta_1(Z_i(\lambda) - x)]^2 K_h(x - Z_i(\lambda))|(Y_i, Z_i)\right)$$

with respect to β_0, β_1 is given by the solution of the following equations

$$\begin{cases} \sum_{i=1}^{n} E\left([Y_i - \beta_0 - \beta_1(Z_i(\lambda) - x)]K_h(Z_i(\lambda) - x)|(Y_i, Z_i)\right) = 0, \\ \sum_{i=1}^{n} E\left([Y_i - \beta_0 - \beta_1(Z_i(\lambda) - x)](Z_i(\lambda) - x)K_h(Z_i(\lambda) - x)|(Y_i, Z_i)\right) = 0. \end{cases} \tag{3.3.2}$$

The choice of kernel function K is not critical in theory, but for the ease of computation, choosing K to be standard normal can bring us extra benefits. In fact, with such a choice, together with the normality of the measurement error, the conditional expectations in (3.3.2) have explicit forms. Note that V_i's are i.i.d. from $N(0, \sigma_u^2)$ and independent of (Z_i, Y_i), routine calculation (see Appendix B) shows that

$$E[K_h(Z(\lambda) - x)|Y, Z] = \phi(x; Z, h^2 + \lambda\sigma_u^2), \tag{3.3.3}$$

$$E[(Z(\lambda) - x)K_h(Z(\lambda) - x)|Y, Z] = \frac{h^2}{h^2 + \lambda\sigma_u^2}(Z - x)\phi(x; Z, h^2 + \lambda\sigma_u^2), \tag{3.3.4}$$

$$\begin{aligned} E[(Z(\lambda) - x)^2 K_h(Z(\lambda) - x)|Y, Z] &= \frac{h^4}{(h^2 + \lambda\sigma_u^2)^2}(Z - x)^2\phi(x; Z, h^2 + \lambda\sigma_u^2) \\ &\quad + \frac{\lambda\sigma_u^2 h^2}{h^2 + \lambda\sigma_u^2}\phi(x; Z, h^2 + \lambda\sigma_u^2), \end{aligned} \tag{3.3.5}$$

here, also throughout this paper, $\phi(x; \mu, \sigma_u^2)$ denotes the normal density function with mean

μ and variance σ_u^2. Denote $A_{nj}(x) = n^{-1}\sum_{i=1}^{n}(Z_i - x)^j\phi(x; Z_i, h^2 + \lambda\sigma_u^2)$ for $j = 0, 1, 2$, and $B_{nl}(x) = n^{-1}\sum_{i=1}^{n} Y_i(Z_i - x)^l\phi(x; Z_i, h^2 + \lambda\sigma_u^2)$ for $l = 0, 1$. Then the solution of (β_0, β_1) of equation (3.3.2), or $(\hat{g}_n(x; \lambda), \hat{g}_n'(x; \lambda))$ has the form of

$$\begin{pmatrix} \hat{g}_n(x;\lambda) \\ \hat{g}_n'(x;\lambda) \end{pmatrix} = \begin{pmatrix} A_{n0}(x) & r(\lambda,h)A_{n1}(x) \\ r(\lambda,h)A_{n1}(x) & r(\lambda,h)[A_{n2}(x) + \lambda\sigma_u^2 A_{n0}(x)] \end{pmatrix}^{-1} \begin{pmatrix} B_{n0}(x) \\ r(\lambda,h)B_{n1}(x) \end{pmatrix}, \tag{3.3.6}$$

where $r(\lambda, h) = h^2/(h^2 + \lambda\sigma_u^2)$.

Note that (3.3.6) itself can be used for extrapolation. However, unlike the estimator $\hat{\beta}(\lambda)$, $\hat{\alpha}(\lambda)$ derived in the example of the linear regression, $\lambda = -1$ cannot be plugged directly into (3.3.6) to get the SIMEX estimator. In fact, when the sample size n gets bigger, the bandwidth h decreases to 0. As a result, when $\lambda = -1$, $h^2 + \lambda\sigma_u^2$ is negative for large sample sizes. As the variance of a normal density function, $h^2 - \sigma_u^2$ should not be negative, which implies the extrapolation step is necessary.

Therefore, we propose the following two-step SIMEX procedure, or more appropriately, the extrapolation (EX) procedure, to find an estimate of the regression function g.

EX Algorithm of The Local Linear Smoother

(1) For each λ from the pre-selected sequence $\lambda = \lambda_1, \ldots, \lambda_K$, calculate $\hat{g}_n(x; \lambda)$;

(2) Identify a trend of the pairs $(\lambda_k, \hat{g}_n(x; \lambda_k))$ and $(\lambda_k, \hat{g}_n'(x; \lambda_k))$, $k = 1, 2, \ldots, K$. Denote the trend as a function $G(x; \lambda)$, respectively. Then, the EX estimator of g and its derivative are defined by $\hat{g}_{\text{EX}}(x) = G(x; -1)$.

Obviously, the above EX algorithm is much more efficient than the classical three-step SIMEX algorithm. Also,it is also easy to see that $\hat{g}_n(x; \lambda)$ from the EX algorithm is not the limit of $\hat{g}_{n,B}(x;\lambda)$ in the SIMEX algorithm as $B \to \infty$. Given the observed data $(Z_i, Y_i)_{i=1}^n$, by the law of large numbers, for a fixed λ, as $B \to \infty$, $\hat{g}_{n,B}(x;\lambda) = B^{-1}\sum_{b=1}^{B} \hat{g}_{n,b}(x;\lambda)$ converges to $\tilde{g}_n(x;\lambda)$ in probability, where

$$\tilde{g}_n(x;\lambda) = \int \frac{S_{2n}(x,\boldsymbol{v})T_{0n}(x,\boldsymbol{v}) - S_{1n}(x,\boldsymbol{v})T_{1n}(x,\boldsymbol{v})}{S_{2n}(x,\boldsymbol{v})S_{0n}(x,\boldsymbol{v}) - S_{1n}^2(x,\boldsymbol{v})}\boldsymbol{\phi}(\boldsymbol{v};0,\lambda\sigma_u^2)d\boldsymbol{v}, \tag{3.3.7}$$

where $\boldsymbol{v} = (v_1, \ldots, v_n)^T$, $\boldsymbol{\phi}(\boldsymbol{v};0,\lambda\sigma_u^2) = \prod_{i=1}^n \phi(v_i;0,\lambda\sigma_u^2)$, and

$$\begin{aligned}
S_{jn}(x,\boldsymbol{v}) &= \frac{1}{n}\sum_{i=1}^n (Z_i + v_i - x)^j K_h(Z_i + v_i - x), \quad j = 0,1,2,\\
T_{ln}(x,\boldsymbol{v}) &= \frac{1}{n}\sum_{i=1}^n (Z_i + v_i - x)^l Y_i K_h(Z_i + v_i - x), \quad l = 0,1.
\end{aligned}$$

However, the estimator $\hat{g}_n(x;\lambda)$ defined in (3.3.6) has the form of

$$\begin{aligned}
\hat{g}_n(x;\lambda) &= \frac{\tilde{S}_{2n}(x)\tilde{T}_{0n}(x) - \tilde{S}_{1n}(x)\tilde{T}_{1n}(x)}{\tilde{S}_{2n}(x)\tilde{S}_{0n}(x) - \tilde{S}_{1n}^2(x)}\\
&= \frac{A_{n2}(x)B_{n0}(x) + \lambda\sigma_u^2 A_{n0}(x)B_{n0}(x) - r(\lambda,h)A_{n1}(x)B_{n1}(x)}{A_{n2}(x)A_{n0}(x) + \lambda\sigma_u^2 A_{n0}^2(x) - r(\lambda,h)A_{n1}^2(x)},
\end{aligned} \tag{3.3.8}$$

where

$$\tilde{S}_{jn}(x) = \int S_{jn}(x,\boldsymbol{v})\boldsymbol{\phi}(\boldsymbol{v};0,\lambda\sigma_u^2)d\boldsymbol{v}, \quad \tilde{T}_{ln}(x) = \int T_{ln}(x,\boldsymbol{v})\boldsymbol{\phi}(\boldsymbol{v};0,\lambda\sigma_u^2)d\boldsymbol{v} \tag{3.3.9}$$

for $j = 0,1,2$ and $l = 0,1$, respectively. Therefore, $\tilde{g}_n(x;\lambda)$ is different from $\hat{g}_n(x;\lambda)$, which indicates that $\hat{g}_n(x;\lambda)$ from the SIMEX algorithm is not the limit of the EX algorithm as $B \to \infty$. In fact, $\hat{g}_n(x;\lambda)$ can be viewed as the limit of $\hat{g}_{n,b}(x;\lambda)$ with $S_{jnb}(x)$ and $T_{lnb}(x)$ replaced by $B^{-1}\sum_{b=1}^B S_{jnb}(x)$ and $B^{-1}\sum_{b=1}^B T_{lnb}(x)$, $j = 0,1,2$, $l = 0,1$, respectively, as $B \to \infty$.

3.4 Asymptotic Theory of EX Algorithm

In this section, we shall investigate the large sample behaviours for the EX algorithm proposed in the previous section. We will show that as $n \to \infty$, $\hat{g}_n(x; \lambda)$ indeed converges to a function of both x and λ, but the latter can approximate the true regression function $g(x)$ as $\lambda \to -1$, thus justifying the effectiveness of extrapolation. The asymptotic joint distribution of $\hat{g}_n(x; \lambda)$ at different λ values, including $\lambda = 0$ which corresponds to the naive estimator, will be also discussed.

The following is a list of regularity conditions we need to justify all the theoretical derivations.

C1. $f_x(x)$, $g(x)$ $\tau^2(x) = E(\varepsilon^2 | X = x)$, $\mu(x) = E(|\varepsilon|^3 | X = x)$ are twice continuously differentiable; also for each x in the support of X, as a function of t, $\eta'(t+x), \eta''(t+x) \in L_2(\phi(t, 0, \sigma_u^2))$, where $\eta = f_x, g, g^2, \tau^2, \tau^4$, μ and μ^2.

C2. The bandwidth h satisfies $h \to 0$, $nh \to \infty$ as $n \to \infty$.

To proceed, for integers $j \geq 0$, we denote

$$f_{j,\lambda}(x) = \int \phi(t; x, (\lambda+1)\sigma_u^2) t^j f_X(t) dt, \quad g_{j,\lambda}(x) = \int t^j g(t) f_X(t) \phi(t, x, (1+\lambda)\sigma_u^2) dt,$$
$$G_{j,\lambda}(x) = \int t^j g^2(t) f_X(t) \phi(t, x, (1+\lambda)\sigma_u^2) dt, \quad H_{j,\lambda}(x) = \int t^j \tau^2(t) f_X(t) \phi(t, x, (1+\lambda)\sigma_u^2) dt.$$

By a routine and tedious calculation, we can show the following result from which the asymptotic bias of $\hat{g}_n(x; \lambda)$ can be derived as $n \to \infty$.

Theorem 4. *Under conditions C1 and C2, for each $\lambda \geq 0$, we have*

$$\frac{E\tilde{S}_{n2}(x) \cdot E\tilde{T}_{n0}(x) - E\tilde{S}_{n1}(x) \cdot E\tilde{T}_{n1}(x)}{E\tilde{S}_{n2}(x) \cdot E\tilde{S}_{n0}(x) - [E\tilde{S}_{n1}(x)]^2} = \frac{g_{0,\lambda}(x)}{f_{0,\lambda}(x)} + h^2 B(x; \lambda) + o(h^2), \qquad (3.4.1)$$

where $B(x; \lambda)$ equals

$$\frac{f_{0,\lambda}(x) g''_{0,\lambda}(x) - f''_{0,\lambda}(x) g_{0,\lambda}(x)}{2 f^2_{0,\lambda}(x)} + \frac{(f_{1,\lambda}(x) - x f_{0,\lambda}(x))(g_{0,\lambda}(x) f_{1,\lambda}(x) - f_{0,\lambda}(x) g_{1,\lambda}(x))}{(\lambda+1)^2 \sigma_u^4 f^3_{0,\lambda}(x)},$$

where $\tilde{S}_{nj}(x)$ and $\tilde{T}_{nl}(x)$ for $j = 0, 1, 2$ and $l = 0, 1$ are defined in (3.3.9).

Note that, as $\lambda \to -1$, $g_{0,\lambda}(x) \to g(x)f_X(x)$, and $f_{0,\lambda}(x) \to f_X(x)$. For $B(x; \lambda)$, we have

$$f_{1,\lambda}(x) - xf_{0,\lambda}(x) = \int (t-x)f_X(t)\phi(t; x, (\lambda+1)\sigma_u^2)dt = (\lambda+1)\sigma_u^2 f_X'(x) + o((\lambda+1)\sigma_u^2),$$

and $g_{1,\lambda}(x) - xg_{0,\lambda}(x)$ can be written as

$$\int (t-x)g(t)f_X(t)\phi(t; x, (\lambda+1)\sigma_u^2)dt = (\lambda+1)\sigma_u^2[gf_X]'(x) + o((\lambda+1)\sigma_u^2)$$

as $\lambda \to -1$. Then we can further show that

$$B(x; \lambda) = \frac{f_{0,\lambda}(x)g''_{0,\lambda}(x) - f''_{0,\lambda}(x)g_{0,\lambda}(x)}{2f^2_{0,\lambda}(x)} + \frac{g_{0,\lambda}(x)(f_X'(x))^2}{f^3_{0,\lambda}(x)} - \frac{f_X'(x)[gf_X]'(x)}{f^2_{0,\lambda}(x)} + o(1), \quad (3.4.2)$$

where $o(1)$ denotes that the corresponding terms converge to 0 as $\lambda \to -1$. Therefore, we have $\lim_{\lambda \to -1} B(x; \lambda) = g''(x)/2$. Thus, from Theorem 4,

$$\lim_{\lambda \to -1} \left[\frac{E\tilde{S}_{n2}(x) \cdot E\tilde{T}_{n0}(x) - E\tilde{S}_{n1}(x) \cdot E\tilde{T}_{n1}(x)}{E\tilde{S}_{n2}(x) \cdot E\tilde{S}_{n0}(x) - [E\tilde{S}_{n1}(x)]^2} \right] = g(x) + \frac{g''(x)h^2}{2} + o(h^2),$$

and this immediately leads to

$$\lim_{\lambda \to -1} \lim_{h \to 0} \left[\frac{E\tilde{S}_{n2}(x) \cdot E\tilde{T}_{n0}(x) - E\tilde{S}_{n1}(x) \cdot E\tilde{T}_{n1}(x)}{E\tilde{S}_{n2}(x) \cdot E\tilde{S}_{n0}(x) - [E\tilde{S}_{n1}(x)]^2} \right] = g(x).$$

To investigate the asymptotic distribution of $\hat{g}_n(x; \lambda)$, denote

$$\begin{aligned}
D_n(x) &= (\tilde{S}_{n2}(x)\tilde{S}_{n0}(x) - \tilde{S}^2_{n1}(x))(E\tilde{S}_{n2}(x)E\tilde{S}_{n0}(x) - (E\tilde{S}_{n1}(x))^2), \\
C_{n0}(x) &= E\tilde{S}_{n2}(x)[E\tilde{S}_{n1}(x)E\tilde{T}_{n1}(x) - E\tilde{S}_{n2}(x)E\tilde{T}_{n0}(x)], \\
C_{n1}(x) &= 2E\tilde{S}_{n1}(x)E\tilde{S}_{n2}(x)E\tilde{T}_{n0}(x) - (E\tilde{S}_{n1}(x))^2E\tilde{T}_{n1}(x) - E\tilde{T}_{n1}(x)E\tilde{S}_{n2}(x)E\tilde{S}_{n0}(x), \\
C_{n2}(x) &= E\tilde{S}_{n1}(x) \cdot [E\tilde{S}_{n0}(x)E\tilde{T}_{n1}(x) - E\tilde{T}_{n0}(x)E\tilde{S}_{n1}(x)], \\
D_{n0}(x) &= E\tilde{S}_{n2}(x)[E\tilde{S}_{n2}(x)E\tilde{S}_{n0}(x) - (E\tilde{S}_{n1}(x))^2],
\end{aligned}$$

$$D_{n1}(x) = E\tilde{S}_{n1}(x)[(E\tilde{S}_{n1}(x))^2 - E\tilde{S}_{n2}(x)E\tilde{S}_{n0}(x)].$$

From Theorem 4, we can write $\hat{g}_n(x)$ as

$$\begin{aligned}\hat{g}_n(x;\lambda) &= \frac{g_{0,\lambda}(x)}{f_{0,\lambda}(x)} + h^2 B(x;\lambda) + o(h^2)\\ &+ D_n^{-1}(x)\Big[\sum_{j=0}^{2} C_{nj}(x)(\tilde{S}_{nj} - E\tilde{S}_{nj}) + \sum_{l=0}^{1} D_{nl}(x)(\tilde{T}_{nl} - E\tilde{T}_{nl})\Big]. \qquad (3.4.3)\end{aligned}$$

Denote

$$\begin{aligned}c_{0\lambda}(x) &= -\frac{g_{0,\lambda}(x)}{f_{0,\lambda}^2(x)}, \quad c_{1\lambda}(x) = \frac{2[f_{1,\lambda}(x) - xf_{0,\lambda}(x)]g_{0,\lambda}(x) - [g_{1,\lambda}(x) - xg_{0,\lambda}(x)]f_{0,\lambda}(x)}{(\lambda+1)\sigma_u^2 f_{0,\lambda}^3(x)},\\ c_{2\lambda}(x) &= \frac{[f_{1,\lambda}(x) - xf_{0,\lambda}(x)][g_{1,\lambda}(x) - xg_{0,\lambda}(x)]f_{0,\lambda}(x) - [f_{1,\lambda}(x) - xf_{0,\lambda}(x)]^2 g_{0,\lambda}(x)}{(\lambda+1)^2\sigma_u^4 f_{0,\lambda}^4(x)},\\ d_{0\lambda}(x) &= \frac{1}{f_{0,\lambda}(x)}, \quad d_{1\lambda}(x) = -\frac{f_{1,\lambda}(x) - xf_{0,\lambda}(x)}{(\lambda+1)\sigma_u^2 f_{0,\lambda}^2(x)}.\end{aligned}$$

Then, from Lemma 9 - Lemma 13 in Appendix B, we can show that, for $\lambda \geq 0$, $D_n^{-1}C_n = c_{j\lambda}(x) + o_p(1)$, for $j = 0, 1, 2$ and $D_n^{-1}D_{nj} = d_{j\lambda}(x) + o_p(1)$ for $j = 0, 1$. We further denote

$$\begin{aligned}\xi_{0\lambda,i}(x) &= \phi(x, Z_i, h^2 + \lambda\sigma_u^2) - E\phi(x, Z, h^2 + \lambda\sigma_u^2),\\ \xi_{1\lambda,i}(x) &= \frac{h^2}{h^2 + \lambda\sigma_u^2}\left[(Z_i - x)\phi(x, Z_i, h^2 + \lambda\sigma_u^2) - E(Z - x)\phi(x, Z, h^2 + \lambda\sigma_u^2)\right],\\ \xi_{2\lambda,i}(x) &= \frac{h^4}{(h^2 + \lambda\sigma_u^2)^2}\left[(Z_i - x)^2\phi(x, Z_i, h^2 + \lambda\sigma_u^2) - E(Z - x)^2\phi(x, Z, h^2 + \lambda\sigma_u^2)\right]\\ &+ \frac{\lambda\sigma_u^2 h^2}{h^2 + \lambda\sigma_u^2}\left[\phi(x, Z_i, h^2 + \lambda\sigma_u^2) - E\phi(x, Z, h^2 + \lambda\sigma_u^2)\right],\\ \eta_{0\lambda,i}(x) &= Y_i\phi(x, Z_i, h^2 + \lambda\sigma_u^2) - EY\phi(x, Z, h^2 + \lambda\sigma_u^2)\\ \eta_{1\lambda,i}(x) &= \frac{h^2}{h^2 + \lambda\sigma_u^2}\left[Y_i(Z_i - x)\phi(x, Z_i, h^2 + \lambda\sigma_u^2) - EY(Z - x)\phi(x, Z, h^2 + \lambda\sigma_u^2)\right].\end{aligned}$$

Then, from (3.4.3), we have

$$\hat{g}_n(x;\lambda) = \frac{g_{0,\lambda}(x)}{f_{0,\lambda}(x)} + h^2 B(x;\lambda) + o(h^2)$$
$$\sum_{j=0}^{2}[c_{j\lambda}(x) + o(1)](\tilde{S}_{nj} - E\tilde{S}_{nj}) + \sum_{k=0}^{1}[d_{k\lambda}(x) + o(1)](\tilde{T}_{nk} - E\tilde{T}_{nk}).$$

Since the terms $o(1)$ in the above expression does not affect the asymptotic distribution of $\hat{g}_n(x;\lambda)$, so we can safely neglect the $o(1)$ terms from the sum, and therefore the two sums can be written as an i.i.d. average $n^{-1}\sum_{i=1}^{n} v_{i\lambda}(x)$, where $v_{i\lambda}(x)$ is defined by

$$c_{0\lambda}(x)\xi_{0\lambda,i}(x) + c_{1\lambda}(x)\xi_{1\lambda,i}(x) + c_{2\lambda}(x)\xi_{2\lambda,i}(x) + d_{0\lambda}(x)\eta_{0\lambda,i}(x) + d_{1\lambda}(x)\eta_{1\lambda,i}(x). \qquad (3.4.4)$$

By verifying the Lyapunov condition, we can show that for each $\lambda > 0$, $n^{-1}\sum_{i=1}^{n} v_{i\lambda}(x)$ is asymptotically normal. This asymptotic normality is summarized in the following theorem.

Theorem 5. *Under conditions C1 and C2, for each $\lambda > 0$,*

$$\sqrt{n}\left\{\hat{g}_n(x;\lambda) - \frac{g_{0,\lambda}(x)}{f_{0,\lambda}(x)} - h^2 B(x;\lambda) + o(h^2)\right\} \Longrightarrow N(0, \Delta_{\lambda,\lambda}(x)),$$

and for $\lambda = 0$,

$$\sqrt{nh}\left\{\hat{g}_n(x;0) - \frac{g_{0,0}(x)}{f_{0,0}(x)} - h^2 B(x;0) + o(h^2)\right\} \Longrightarrow N(0, \Delta_{0,0}(x)),$$

where

$$\Delta_{\lambda,\lambda}(x) = c_{0\lambda}^2\left[\frac{f_{0,\lambda/2}(x)}{2\sqrt{\pi\lambda\sigma_u^2}} - f_{0,\lambda}^2(x)\right] + d_{0\lambda}^2\left[\frac{G_{0,\lambda/2}(x) + H_{0,\lambda/2}(x)}{2\sqrt{\pi\lambda\sigma_u^2}} - g_{0,\lambda}^2\right]$$
$$+2c_{0\lambda}d_{0\lambda}\left[\frac{g_{0,\lambda/2}(x)}{2\sqrt{\pi\lambda\sigma_u^2}} - g_{0,\lambda}(x)f_{0,\lambda}(x)\right]$$

and

$$\Delta_{0,0}(x) = \frac{1}{2\sqrt{\pi}}\left[\frac{G_{00}(x)+H_{00}(x)}{f_{00}^2(x)} - \frac{g_{00}^2(x)}{f_{00}^3(x)}\right].$$

Note that when $\sigma_u^2 = 0$, that is, no measurement error in X, then one can easily see that $\Delta_{0,0}(x) = \tau^2(x)/(2\sqrt{\pi}f_X(x))$, which is exactly the asymptotic variance in local linear estimator of the regression function in the error-free cases. The theorem below states the asymptotic joint normality of $[\hat{g}_n(x;0), \hat{g}_n(x;\lambda_1), \cdots, \hat{g}_n(x;\lambda_K)]'$.

Theorem 6. *Under conditions C1 and C2, for $0 < \lambda_1 < \cdots < \lambda_K < \infty$,*

$$\begin{pmatrix} \sqrt{nh} & 0 & \cdots & 0 \\ 0 & \sqrt{n} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{n} \end{pmatrix} \begin{pmatrix} \hat{g}_n(x;0) - g_{0,0}(x)/f_{0,0}(x) - h^2B(x;0) + o(h^2) \\ \hat{g}_n(x;\lambda_1) - g_{0,\lambda_1}(x)/f_{0,\lambda_1}(x) - h^2B(x;\lambda_1) + o(h^2) \\ \vdots \\ \hat{g}_n(x;\lambda_K) - g_{0,\lambda_K}(x)/f_{0,\lambda_K}(x) - h^2B(x;\lambda_K) + o(h^2) \end{pmatrix}$$

$$\Longrightarrow N(0, \Delta(x)),$$

where $B(x;\lambda)$ is defined in (3.4.2),

$$\Delta(x) = \begin{pmatrix} \Delta_{0,0}(x) & 0 & 0 & \cdots & 0 \\ 0 & \Delta_{\lambda_1,\lambda_1}(x) & \Delta_{\lambda_1\lambda_2}(x) & \cdots & \Delta_{\lambda_1\lambda_K}(x) \\ 0 & \Delta_{\lambda_1\lambda_2}(x) & \Delta_{\lambda_2,\lambda_2}(x) & \cdots & \Delta_{\lambda_2\lambda_K}(x) \\ \vdots & \Delta_{\lambda_1\lambda_K}(x) & \Delta_{\lambda_2\lambda_K}(x) & \cdots & \Delta_{\lambda_K,\lambda_K}(x) \end{pmatrix},$$

and $\Delta_{\lambda_i\lambda_j}(x)$, $i = 0, 1, \ldots, K$, $j = 1, 2, \ldots, K$, are given by

$$\begin{aligned} &\frac{c_{0\lambda_i}(x)c_{0\lambda_j}(x)}{\sqrt{2\pi(\lambda_i+\lambda_j)\sigma_u^2}}\int \phi\left(t, x, \left(\frac{\lambda_i\lambda_j}{\lambda_i+\lambda_j}+1\right)\sigma_u^2\right) f_X(t)dt - f_{0,\lambda_i}(x)f_{0,\lambda_j}(x) \\ &+\frac{c_{0\lambda_i}(x)d_{0\lambda_j}(x)}{\sqrt{2\pi(\lambda_i+\lambda_j)\sigma_u^2}}\int g(t)\phi\left(t, x, \left(\frac{\lambda_i\lambda_j}{\lambda_i+\lambda_j}+1\right)\sigma_u^2\right) f_X(t)dt - f_{0,\lambda_i}(x)g_{0,\lambda_j}(x) \\ &+\frac{c_{0\lambda_j}(x)d_{0\lambda_i}(x)}{\sqrt{2\pi(\lambda_i+\lambda_j)\sigma_u^2}}\int g(t)\phi\left(t, x, \left(\frac{\lambda_i\lambda_j}{\lambda_i+\lambda_j}+1\right)\sigma_u^2\right) f_X(t)dt - f_{0,\lambda_j}(x)g_{0,\lambda_i}(x) \end{aligned}$$

$$+\frac{d_{0\lambda_i}(x)d_{0\lambda_j}(x)}{\sqrt{2\pi(\lambda_i+\lambda_j)\sigma_u^2}}\int g^2(t)\phi\left(t,x,\left(\frac{\lambda_i\lambda_j}{\lambda_i+\lambda_j}+1\right)\sigma_u^2\right)f_X(t)dt-g_{0,\lambda_i}(x)g_{0,\lambda_j}(x).$$

The proof of the joint normality is a straightforward application of the multivariate CLT on the following random vector

$$\begin{pmatrix}\sqrt{nh}\left[\hat{g}_n(x;0)-g_{0,0}(x)/f_{0,0}(x)-h^2B(x;0)+o(h^2)\right]\\ \sqrt{n}\left[\hat{g}_n(x;\lambda_1)-g_{0,\lambda_1}(x)/f_{0,\lambda_1}(x)-h^2B(x;\lambda_1)+o(h^2)\right]\\ \vdots\\ \sqrt{n}\left[\hat{g}_n(x;\lambda_K)-g_{0,\lambda_K}(x)/f_{0,\lambda_K}(x)-h^2B(x;\lambda_K)+o(h^2)\right]\end{pmatrix}=\frac{1}{\sqrt{n}}\begin{pmatrix}\sqrt{h}\sum_{i=1}^n v_{i0}(x)\\ \sum_{i=1}^n v_{i\lambda_1}(x)\\ \vdots\\ \sum_{i=1}^n v_{i\lambda_K}(x)\end{pmatrix}.$$

For the sake of brevity, the proof will be omitted. In addition to the condition C2, if we further assume that $nh^4\to 0$, then the asymptotic bias can be removed.

3.5 Extrapolation Function

From the discussion in the previous section, the extrapolation function can be derived from $g_{0,\lambda}(x)/f_{0,\lambda}(x)$. From the definitions of $g_{0,\lambda}$ and $f_{0,\lambda}(x)$, we know that

$$\Gamma(\lambda):=\frac{g_{0,\lambda}(x)}{f_{0,\lambda}(x)}=\frac{\int g(t)f_X(t)\phi(t;x,(\lambda+1)\sigma_u^2)dt}{\int f_X(t)\phi(t;x,(\lambda+1)\sigma_u^2)dt}. \tag{3.5.1}$$

As a function of λ, $\Gamma(\lambda)$ does not have a tractable form, and some approximation is needed for extrapolating. By change of variable, we have

$$\int g(t)f_X(t)\phi(t;x,(\lambda+1)\sigma_u^2)dt=\int g(x+\sqrt{\lambda+1}\sigma_u v)f_X(x+\sqrt{\lambda+1}\sigma_u v)\phi(v)dv.$$

Denote $\alpha=(\lambda+1)\sigma_u^2$, and assume that g and f_X are four times continuously differentiable. Then we have

$$\int g(t)f_X(t)\phi(t;x,(\lambda+1)\sigma_u^2)dt=g(x)f_X(x)+\frac{[f_X(x)g(x)]''}{2}\alpha+\frac{[f_X(x)g(x)]^{(4)}}{4!}\alpha^2+o(\alpha^2),$$

where $o(\cdot)$ is understood as a negligible quantity when $\lambda \to -1$. Similarly, we have

$$\int f_X(t)\phi(x; t, (\lambda+1)\sigma_u^2)dt = f_X(x) + \frac{f_X''(x)}{2}\alpha + \frac{f_X^{(4)}(x)}{4!}\alpha^2 + o(\alpha^2).$$

Therefore, after neglecting the $o(\lambda+1)$ term, from (3.5.1), we obtain

$$\Gamma(\lambda) \approx \frac{g(x)f_X(x) + \sigma_u^2(\lambda+1)[f_X(x)g(x)]''/2}{f_X(x) + \sigma_u^2(\lambda+1)f_X''(x)/2}.$$

It is easy to see that the right hand side approaches $g(x)$ as $\lambda \to -1$, and indeed, for fixed x-value, it has the form of $a + b/(c+\lambda)$, the nonlinear extrapolation function often used in the classical SIMEX estimation procedure. If we further apply the approximation

$$\frac{1}{f_X(x) + \sigma_u^2(\lambda+1)f_X''(x)/2} = \frac{1}{f_X(x)}\left[1 - \frac{\sigma_u^2(\lambda+1)f_X''(x)}{2f_X(x)} + o((\lambda+1))\right]$$

or the approximation with higher order expansions, then we can obtain the commonly used quadratic extrapolation function $a + b\lambda + c\lambda^2$ and the polynomial extrapolation functions.

Almost all literature involving the classical SIMEX method, mostly in the parametric setups, assumes that the true extrapolation function has a known nonlinear form when discussing the asymptotic distributions of the SIMEX estimators. However, based on the above discussion, the true extrapolation function is never known. To see this point clearly, we further assume that $X \sim N(0, \sigma_x^2)$. Then from (3.5.1), for any $x \in \mathbb{R}$,

$$\Gamma(\lambda) = \int g(t)\phi\left(t, \frac{x\sigma_x^2}{(\lambda+1)\sigma_u^2 + \sigma_x^2}, \frac{(\lambda+1)\sigma_u^2\sigma_x^2}{(\lambda+1)\sigma_u^2 + \sigma_x^2}\right)dt.$$

Since the normal distribution family is complete, so the above expression implies that $\Gamma(\lambda)$ and $g(t)$ are uniquely determined by each other. Since g is unknown, so neither is Γ. This discouraging finding really invalidates all the potential theoretical developments based on known extrapolation functions.

3.6 Numerical Study

In this section, we conduct some simulation studies to evaluate the finite sample performance of the proposed SIMEX procedure. We also analyze a dataset from the National Health and Nutrition Examination Survey (NHANES) to illustrate the application of the proposed estimation procedure.

3.6.1 Simulation Study

In this simulation study, the simulated data are generated from the regression model $Y = g(X) + \varepsilon, Z = X + U$, where X has a standard normal distribution, U is generated from $N(0, \sigma_u^2)$ and ε is generated from a standard normal distribution. Three choices of regression function $g(x)$ were considered, namely $g(x) = x^2$, $\exp(x)$ and $x\sin(x)$. To see the effect of the measurement error variance on the resulting estimate, we choose $\sigma_u^2 = 0.1$ and 0.25. The sample sizes are chosen to be $n = 100, 200, 500$. In each scenario, the estimates are calculated for 200 equally spaced x-values x_j, $j = 1, 2, \ldots, 200$, are chosen from $[-3, 3]$. To implement the extrapolation step, the grid of λ is taken from 0 to 2 separated by 0.2. The mean squared errors (MSE) are used to evaluate the finite sample performance of the proposed SIMEX procedure. The bandwidth h is chosen to be $n^{-1/5}$, a theoretical optimal order when estimating the regression function based on the error-free data. To get a stable result, all simulations were performed for 10 independent datasets, and the average of the 10 estimates was taken to be the final estimate at each of 200 x-values, and the MSE defined by $200^{-1}\sum_{j=1}^{200}[\hat{g}_n(x_j) - g(x_j)]^2$ is used for evaluate the finite sample performance of the proposed EX estimate. For comparison, we also apply the classical SIMEX algorithm and the naive method to estimate these three regression functions with $B = 50, 100$. Besides the report on the MSEs from the three algorithms, we also record the computation time in seconds from each procedure to evaluate the algorithm efficiency. All the simulations were conducted on a desktop computer with Windows system running on Intel(R) Core(TM) i7-10700 CPU of 2.90GHz, and 16GB RAM. The quadratic extrapolation function is used for both the EX and SIMEX procedures. The simulation results are summarized in Tables 3.1 - 3.3. In all the tables, we use the EX to denote the proposed Extrapolation algorithm, SIMEX to denote the classical SIMEX method, and Naive for naive method.

Table 3.1: $g(x) = x\sin(x)$, $X \sim N(0, 1)$

σ_u^2	Method		$n = 100$		$n = 200$		$n = 500$	
			MSE	Time(s)	MSE	Time(s)	MSE	Time(s)
0.1	SIMEX	$B = 50$	0.066	71.001	0.065	122.568	0.023	285.839
		$B = 100$	0.142	139.862	0.137	243.531	0.032	569.185
	EX		0.367	2.264	0.079	3.120	0.051	5.853
	Naive		0.197	0.180	0.101	0.274	0.066	0.569
0.25	SIMEX	$B = 50$	0.075	71.053	0.047	123.463	0.077	287.537
		$B = 100$	0.084	141.015	0.138	245.227	0.067	572.512
	EX		0.335	2.202	0.118	2.948	0.065	5.334
	Naive		0.221	0.176	0.139	0.273	0.119	0.574

Table 3.2: $g(x) = x^2$, $X \sim N(0, 1)$

σ_u^2	Method		$n = 100$		$n = 200$		$n = 500$	
			MSE	Time(s)	MSE	Time(s)	MSE	Time(s)
0.1	SIMEX	$B = 50$	0.420	70.813	0.346	122.444	0.055	285.444
		$B = 100$	0.165	139.648	0.192	243.118	0.154	568.681
	EX		0.688	2.263	0.201	3.129	0.067	5.884
	Naive		0.318	0.179	0.450	0.270	0.385	0.572
0.25	SIMEX	$B = 50$	0.996	71.103	0.336	123.314	0.164	287.267
		$B = 100$	0.711	140.884	0.070	245.532	0.196	573.299
	EX		0.069	2.176	0.372	2.949	0.029	5.340
	Naive		1.820	0.180	2.282	0.274	1.456	0.573

The simulation results clearly show that the proposed EX algorithm is more efficient than the classical SIMEX method in terms of computational speed. The finite sample performance of both methods SIMEX and EX, as measured by the MSE, are comparable. When sample sizes get larger, and the measurement error variances get smaller, both procedures performs better, as expected. The advantage of using EX or SIMEX over the naive method may not be obvious when the sample size or the noise level σ_u^2 is small, but both methods outperform

Table 3.3: $g(x) = \exp(x)$, $X \sim N(0, 1)$

σ_u^2	Method		$n = 100$		$n = 200$		$n = 500$	
			MSE	Time(s)	MSE	Time(s)	MSE	Time(s)
0.1	SIMEX	$B = 50$	1.542	70.906	1.654	122.614	0.060	285.777
		$B = 100$	0.351	139.888	0.069	243.476	0.252	568.903
	EX		1.144	2.250	0.125	3.157	0.149	5.826
	Naive		0.637	0.175	1.282	0.273	0.798	0.557
0.25	SIMEX	$B = 50$	2.241	71.761	1.707	123.365	0.898	287.329
		$B = 100$	0.544	140.633	0.150	245.134	0.217	572.378
	EX		0.208	2.176	0.436	2.950	0.070	5.342
	Naive		3.772	0.176	3.785	0.274	2.749	0.570

the naive method when either the sample size or the noise level is increased. It is well known in measurement error literature that the performance of the estimation procedure heavily depends on the signal to noise ratio, or the ratio of σ_x^2 and σ_u^2. The signal to noise ratios in the previous simulation studies are 10 and 4. We also conducted some simulation studies with signal to noise ratio changed to 40 and 16 This resulted in improved performance of all three methods, with the naive method sometimes providing better results than the SIMEX and EX methods, which was not unexpected, since such high signal to noise ratios imply the effect of measurement error is nearly negligible.

As mentioned in the beginning, we used the average of the estimates from 10 independent data sets as the final estimate of the regression function. For illustration purposes, in Figure 3.1 to Figure 3.9, for each simulation setup, we present the fitted the regression curves for $n = 200$, $\sigma_u^2 = 0.25$, and $B = 50$ for SIMEX, from all three methods, with the true regression function as the reference. For completeness, in each figure, the four small plots show the fitted regression curves from four different data sets, and the large plot shows the fitted regression curve based on the averages.

Figure 3.1: Naive Estimate: $g(x) = x\sin(x)$

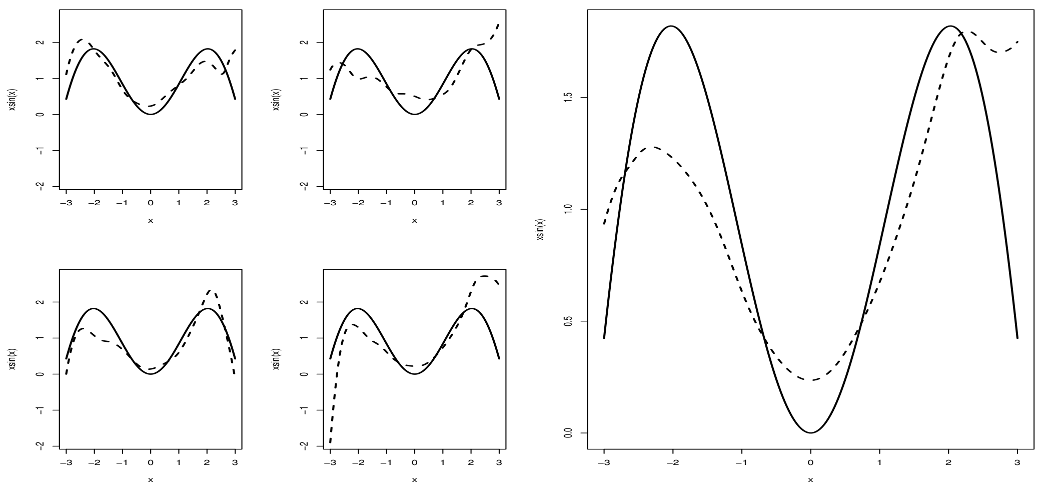
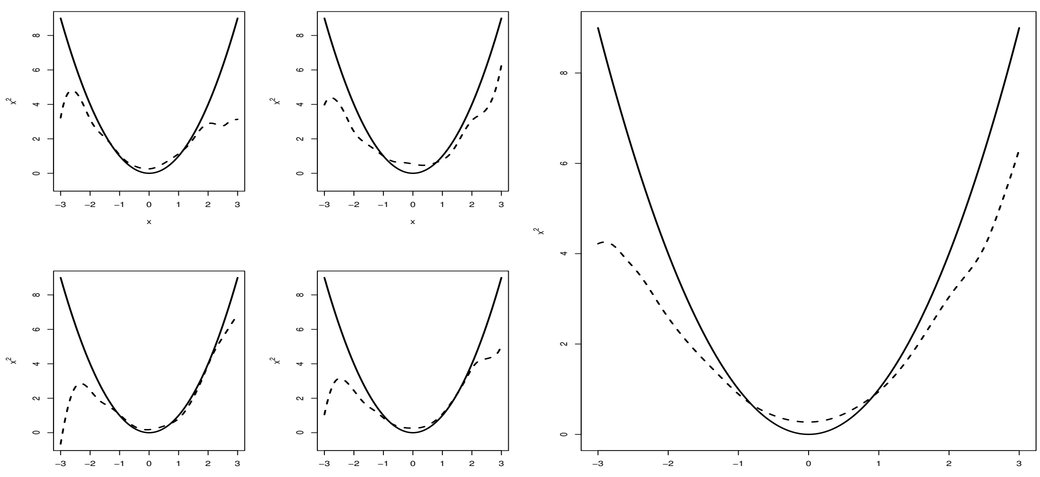


Figure 3.2: Naive Estimate: $g(x) = x^2$



3.6.2 Real Data Application: NHANES Data Set

To determine the relationship between the serum 25-hydroxyvitamin D (25(OH)D) and the long term vitamin D average intake, Curley (2017) analyzed a data set from the National Health and Nutrition Examination Survey (NHANES), and used a nonlinear function for modeling the regression mean of 25(OH)D on the long term vitamin D average intake. In this section, we apply the proposed estimation procedure on a subset of the 2009-2010 NHANES study. The selected data set contains dietary records of 806 Mexican-American females. The long term vitamin D average intake (X) is not measured directly, instead, two

Figure 3.3: Naive Estimate: $g(x) = \exp(x)$

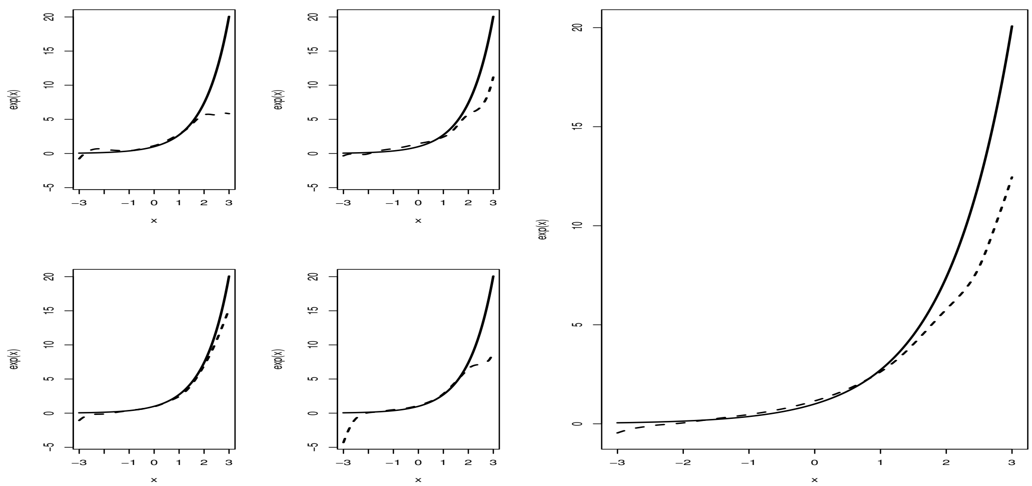
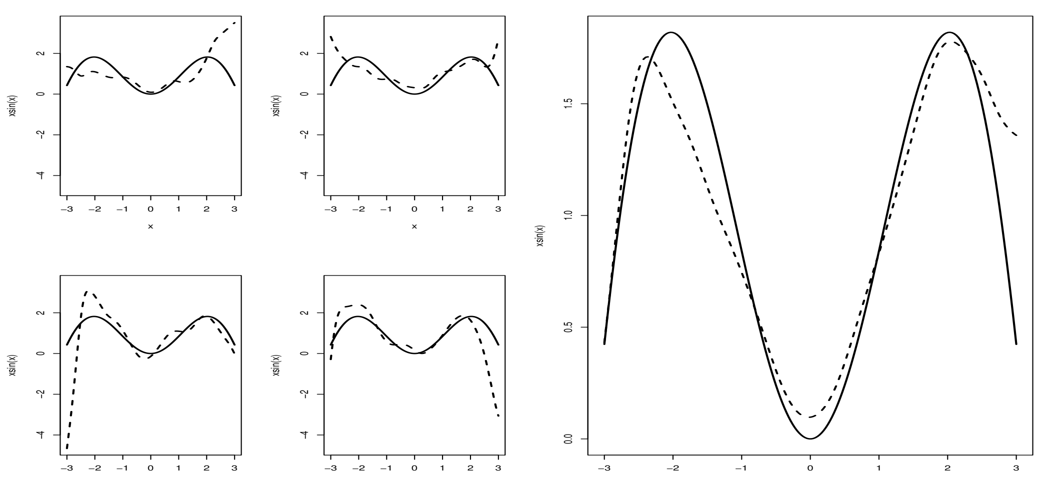


Figure 3.4: SIMEX Estimate: $g(x) = x\sin(x)$



independent daily observations of vitamin D intake are collected. Let W_{ji} be the vitamin D intake from the i-th subject on the j-th time, and we assume that the additive structures $W_{ji} = X_i + U_{ji}$ hold for all $i = 1, 2, \ldots, 806$, $j = 1, 2$. We use $W_i = (W_{1i} + W_{2i})/2$ to represent the observed vitamin intake, and by assuming that U_{1i} and U_{2i} are independently and identically distributed, we can estimate the standard deviation of the measurement error U by the sample standard deviation of the differences $(W_{1i} - W_{2i})/2$, $i = 1, 2, \ldots, n$. As in Curley (2017), we also apply a square root transformation on the 25(OH)D which results in a more symmetric structure, but the Shapiro normal test reports a p-value of 0.04, indicating

Figure 3.5: SIMEX Estimate: $g(x) = x^2$

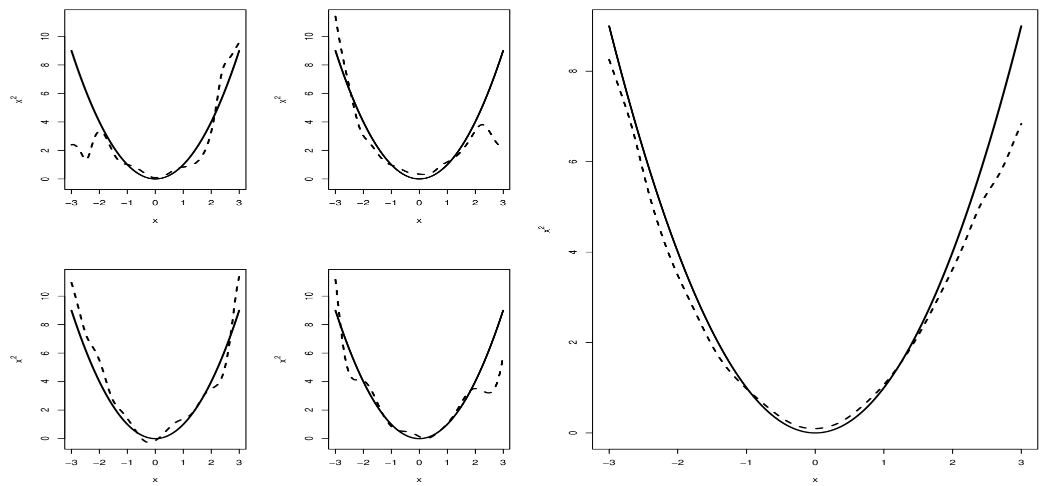
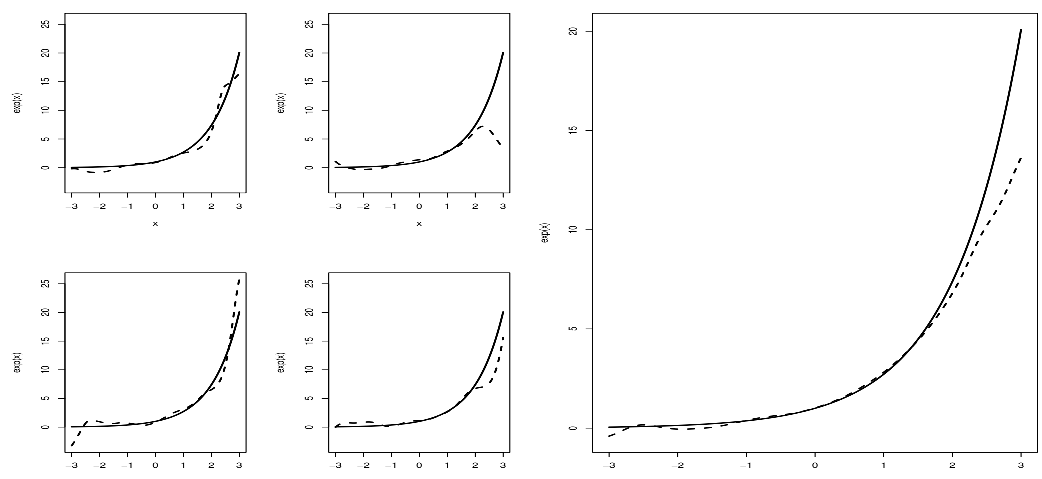


Figure 3.6: SIMEX Estimate: $g(x) = \exp(x)$



that the transformed 25(OH)D values, denoted as Y, is still not normal.

We adopt the local linear estimator to fit the regression function of Y against X to capture the mean regression function using the Naive, SIMEX ($B = 200$) and the proposed EX methods. Three fitted regression functions with the bandwidth $h = n^{-1/5}$, together with the scatter plots of Y against W, are plotted in Figure 3.10. In Figure 3.10, the solid line is the fitted EX regression function, the dashed line is the fitted regression function using the classical SIMEX, and the dotted line is the fitted regression curve using the naive method. Clearly the naive estimator captures the central structure of the raw data, as expected.

Figure 3.7: EX Estimate: $g(x) = x\sin(x)$

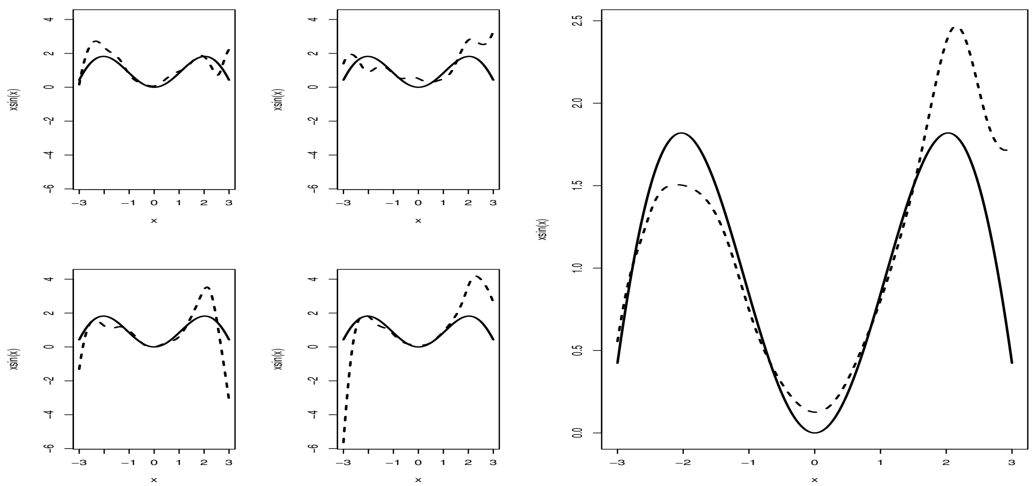
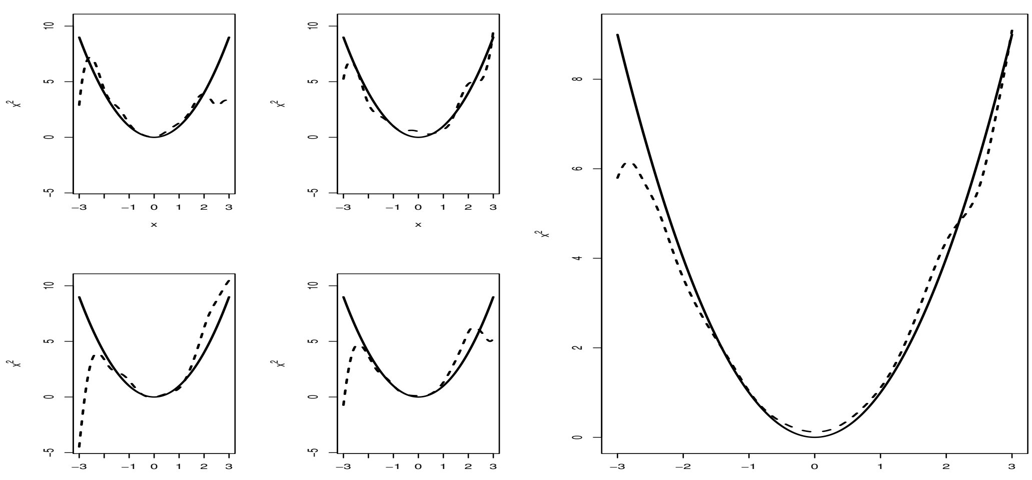
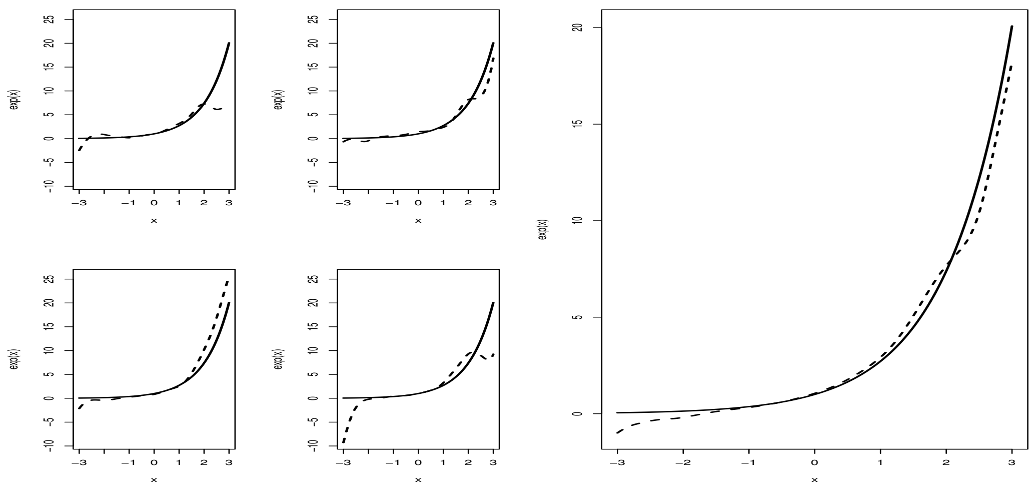


Figure 3.8: EX Estimate: $g(x) = x^2$



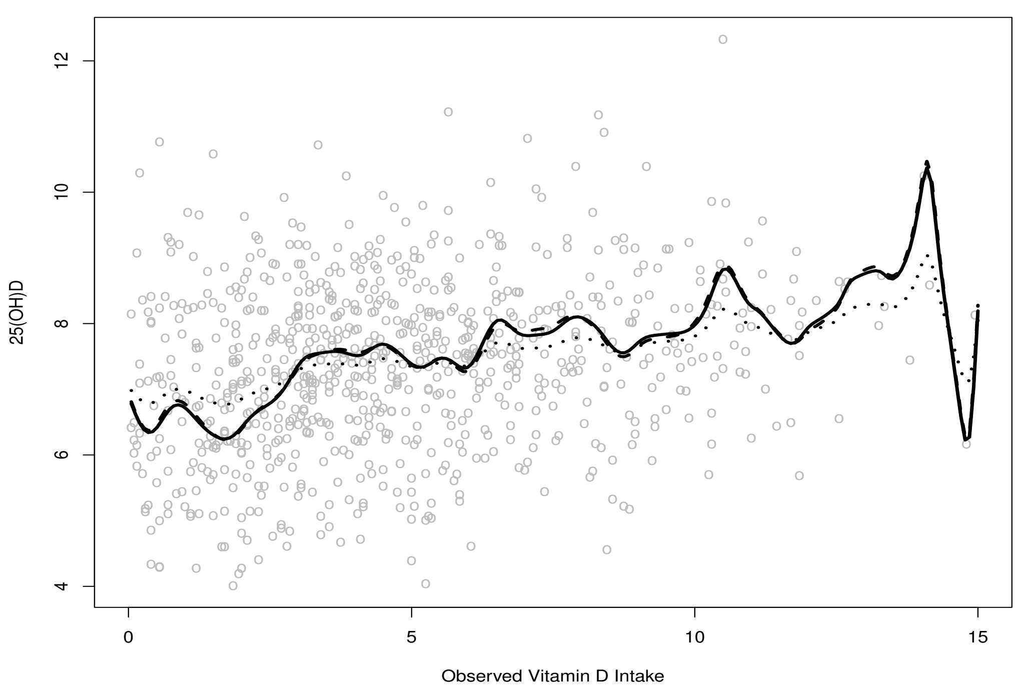
The fitted regression function from the classical SIMEX nearly overlaps the proposed EX estimator. Compared with the naive regression, the SIMEX and the EX procedures provide relatively conservative fitted 25(OH)D values when the vitamin D intake values are small, which might be interpreted as an evidence of the subjects under-reporting their vitamin D intakes. Because fewer data points on the upper end, so we truncated the graph when the observed vitamin D intake is bigger than 15, therefore more caution should be paid when interpreting the trend on the right. More scientific explanations from the analysis need to consult with experts on nutrition studies. The computation times for each of the three

Figure 3.9: EX Estimate: $g(x) = \exp(x)$



methods are, 0.209 seconds for Naive, 0.839 seconds for the EX, and 209.58 seconds for the classical SIMEX. Again, one can see that the proposed EX method is more efficient than the classical SIMEX.

Figure 3.10: Naive, SIMEX and EX estimates



3.7 Discussion

Instead of taking the conditional expectation of the estimator based on the pseudo-data or following the three steps in the classical SIMEX algorithm, the proposed EX method applies the conditional expectation directly to the target function to be optimized based on the pseudo-data, thus successfully bypassing the simulation step. Both the simulation studies and the real data applications indicate the EX algorithm is more effective than the classical SIMEX, as evidenced by less computation time and smaller MSEs. In Section 3.3, we discussed the main difference between $\hat{g}_n(x;\lambda)$ and $\tilde{g}_n(x;\lambda)$, but a more detailed comparison can be made by a heuristic argument as shown below. Define $\boldsymbol{s} = (s_0, s_1, s_2, t_0, t_1)^T$, and a function

$$F(\boldsymbol{s}) = F(s_0, s_1, s_2, t_0, t_1) = \frac{s_2 t_0 - s_1 t_1}{s_2 s_0 - s_1^2},$$

then a Taylor expansion at $\tilde{\boldsymbol{s}} = (\tilde{s}_0, \tilde{s}_1, \tilde{s}_2, \tilde{t}_0, \tilde{t}_1)^T$ up to order 2 leads to

$$F(\boldsymbol{s}) \doteq F(\tilde{\boldsymbol{s}}) + \left.\frac{\partial F(\boldsymbol{s})}{\partial \boldsymbol{s}}\right|_{\boldsymbol{s}=\tilde{\boldsymbol{s}}}^{T} (\boldsymbol{s} - \tilde{\boldsymbol{s}}) + \frac{1}{2}(\boldsymbol{s} - \tilde{\boldsymbol{s}})^T \left.\frac{\partial^2 F(\boldsymbol{s})}{\partial \boldsymbol{s} \partial \boldsymbol{s}^T}\right|_{\boldsymbol{s}=\tilde{\boldsymbol{s}}} (\boldsymbol{s} - \tilde{\boldsymbol{s}}).$$

Let $s_j = S_{jn}(x, \boldsymbol{v})$, $\tilde{s}_j = \tilde{S}_{jn}(x)$ for $j = 0, 1, 2$, and $t_j = T_{jn}(x, \boldsymbol{v})$, $\tilde{t}_j = \tilde{T}_{jn}(x)$ for $j = 0, 1$. Then it is easy to see that $\tilde{g}_n(x;\lambda) = E[F(\boldsymbol{s})|\boldsymbol{D}]$, and $\hat{g}_n(x;\lambda) = F(\tilde{\boldsymbol{s}})$. Therefore, from the above Taylor expansion, we can see that $\tilde{g}_n(x;\lambda) - \hat{g}_n(x;\lambda)$ approximately equals

$$E\left((\boldsymbol{s} - \tilde{\boldsymbol{s}})^T \left.\frac{\partial^2 F(\boldsymbol{s})}{\partial \boldsymbol{s} \partial \boldsymbol{s}^T}\right|_{\boldsymbol{s}=\tilde{\boldsymbol{s}}} (\boldsymbol{s} - \tilde{\boldsymbol{s}}) \middle| \boldsymbol{D} \right) = \operatorname{trace}\left(\left.\frac{\partial^2 F(\boldsymbol{s})}{\partial \boldsymbol{s} \partial \boldsymbol{s}^T}\right|_{\boldsymbol{s}=\tilde{\boldsymbol{s}}} E\left((\boldsymbol{s} - \tilde{\boldsymbol{s}})(\boldsymbol{s} - \tilde{\boldsymbol{s}})^T|\boldsymbol{D}\right) \right).$$

Note that the matrix $E\left((\boldsymbol{s} - \tilde{\boldsymbol{s}})(\boldsymbol{s} - \tilde{\boldsymbol{s}})^T|\boldsymbol{D}\right)$ is nonnegative definite, so it is sufficient to consider the expectation of each entry only in the matrix to determine its order. We can show that all 25 terms are of the order $O(1/nh)$ for $\lambda \geq 0$. This implies that, for $\lambda > 0$, the estimators $\hat{g}_n(x;\lambda)$ and $\tilde{g}_n(x;\lambda)$ are equivalent, in the sense of having the same asymptotic distribution, if $nh^4 \to 0$, and for $\lambda = 0$, they are equivalent if $nh^5 \to 0$. All the necessary computations supporting these claims can be found in the supplement materials.

3.7.1 On the extrapolation function

As we discussed in Section 5, theoretically it is impossible to specify the true form of the extrapolation function in the nonparametric regression setups. However, if the regression function g has a parametric form, then according to the above discussion, we can indeed nail down the extrapolation function. To see this, consider the power function x^p with some $p \geq 1$, with X still assumed to be $N(0, \sigma_x^2)$. Then some algebra leads to

$$\int t^p \phi\left(t, \frac{x\sigma_x^2}{(\lambda+1)\sigma_u^2+\sigma_x^2}, \frac{(\lambda+1)\sigma_u^2\sigma_x^2}{(\lambda+1)\sigma_u^2+\sigma_x^2}\right) dt = \sum_{j=0}^{[p/2]} \binom{p}{2j} (2j-1)!!x^{p-2j} \frac{\sigma_x^{2p}\sigma_u^{4j}(\lambda+1)^{2j}}{[(\lambda+1)\sigma_u^2+\sigma_x^2]^p}.$$

This implies that for a polynomial regression function g of order p, if X is normal, then the extrapolation function can be taken as a polynomial function of order p. Without loss of generality, assume that $H(\lambda) = s^T(\lambda)\boldsymbol{\alpha}$, where $s(\lambda) = (1, \lambda, \lambda^2, \cdots, \lambda^p)^T$. Then $\boldsymbol{\alpha}$ can be estimated by the minimizer of $L(\boldsymbol{\alpha}) = \sum_{j=0}^{K}[\hat{g}_{\lambda_j}(x) - s^T(\lambda_j)\boldsymbol{\alpha}]^2$. In fact, the minimizer $\hat{\boldsymbol{\alpha}} = \left[\sum_{j=0}^{K} s(\lambda_j)s^T(\lambda_j)\right]^{-1} \sum_{j=0}^{K} \hat{g}_{\lambda_j}(x)s(\lambda_j)$. Similar to Carroll et al. (1999), we have, for $nh^5 \to 0$, $\sqrt{nh}(\hat{\alpha}-\alpha)$ is asymptotically normal with mean $\mu(x,h) = h^2B(x,0)\left[\sum_{j=0}^{K} s(\lambda_j)s^T(\lambda_j)\right]^{-1} s(0)$, and covariance matrix

$$\tau^2(x) = \Delta_0(x)\left[\sum_{j=0}^{K} s(\lambda_j)s^T(\lambda_j)\right]^{-1} s(0)s^T(0)\left[\sum_{j=0}^{K} s(\lambda_j)s^T(\lambda_j)\right]^{-1}.$$

Thus, for the SIMEX estimator $s^T(-1)\hat{\boldsymbol{\alpha}}$, we have $\sqrt{nh}(s^T(-1)\hat{\boldsymbol{\alpha}} - g(x) - s^T(-1)\mu(x,h))$ converges to $N(0, \tau^2(x))$ in distribution. Of course, the discussion only has some theoretical significance. If we knew in advance that g has a parametric form, we would not estimate it using the nonparametric methods.

Bibliography

R.J. Carroll, D. Rupper, C.M. Crainiceanu, and L.A. Stefanski. Measurement error in nonlinear models: a modern perspective. Chapman and Hall/CRC, 2 edition, 2006.

J. R. Cook and L. A. Stefanski. Simulation-extrapolation estimation in parametric measurement error models. Journal of the American Statistical Association, 89(428):1314–1328, 1994. ISSN 0162-1459.

Hira L. Koul and Weixing Song. Simulation extrapolation estimation in parametric models with laplace measurement error. Electronic Journal of Statistics, 8:1973–1995, 2014.

Raymond J. Carroll, Jeffrey D. Maca, and David Ruppert. Nonparametric regression in the presence of measurement error. Biometrika, 86(3):541–554, 1999.

J. Fan and Y. K. Truong. Nonparametric regression with errors in variables. Annals of Statistics, 21:1900–1925, 1993.

Kanwal Ayub, Weixing Song, and Jianhong Shi. Extrapolation estimation in parametric regression models with measurement error. Computational Statistics & Data Analysis, 172:107478, 2022. ISSN 0167-9473.

L. A. Stefanski and J. R. Cook. Simulation-extrapolation: The measurement error jackknife. Journal of the American Statistical Association, 90(432):1247–1256, 1995. ISSN 0162-1459.

R. J. Carroll, H. Kuchenhoff, F. Lombard, and L. A. Stefanski. Asymptotics for the simex estimator in nonlinear measurement error models. Journal of the American Statistical Association, 91(433):9, 1996.

Jie Guo and Tong Li. Poisson regression models with errors-in-variables: Implication and treatment. Journal of Statistical Planning and Inference, 104:391–401, 06 2002.

K. Chen, S. Guo, Y. Lin, and Z. Ying. Least absolute relative error estimation. Journal of the American Statistical Association, 105(491):1104–1112, 2010.

K. Chen, Y. Lin, Z. Wang, and Z. Ying. Least product relative error estimation. Journal of Multivariate Analysis, 144:91–98, 2016.

Long Feng, Changliang Zou, and Zhaojun Wang. Local walsh-average regression. 106:36–48, apr 2012. ISSN 0047-259X. doi: 10.1016/j.jmva.2011.12.003. URL https://doi.org/10.1016/j.jmva.2011.12.003.

T. S. Ferguson. A couse in large sample theory. Chapman and Hall/CRC, 1 edition, 1996.

Hua Liang, Wolfgang Härdle, and Raymond J. Carroll. Estimation in a semiparametric partially linear errors-in-variables model. The Annals of Statistics, 27(5):1519 – 1535, 1999.

Brenna Curley. Nonlinear models with measurement error: Application to vitamin D. PhD thesis, Iowa State University, 2017.

Tsuyoshi Nakamura. Corrected score function for errors-in-variables models: Methodology and application to generalized linear models. Biometrika, 77(1):127–137, 1990. ISSN 00063444.

Ranjeeta Mallick, K Fung, and D Krewski. Adjusting for measurement error in the cox proportional hazards regression model. Journal of cancer epidemiology and prevention, 7 (4):155–164, 2002.

Varadan Sevilimedu, Lili Yu, Hani Samawi, and Haresh Rochani. Application of the misclassification simulation extrapolation procedure to log-logistic accelerated failure time models in survival analysis. Journal of Statistical Theory and Practice, 13(1):1–16, 2019.

WR Gould, LA Stefanski, and KH Pollock. Use of simulation–extrapolation estimation in catch–effort analyses. Canadian Journal of Fisheries and Aquatic Sciences, 56(7):1234–1240, 1999.

Wen-Han Hwang and Steve YH Huang. Estimation in capture-recapture models when covariates are subject to measurement errors. Biometrics, 59(4):1113–1122, 2003.

Jakub Stoklosa, Peter Dann, Richard M. Huggins, and Wen-Han Hwang. Estimation of survival and capture probabilities in open population capture–recapture models when covariates are subject to measurement error. Computational Statistics & Data Analysis, 96:74–86, 2016.

Xihong Lin and Raymond J Carroll. Simex variance component tests in generalized linear mixed measurement error models. Biometrics, 55(2):613–619, 1999.

James W Hardin, Henrik Schmiediche, and Raymond J Carroll. The simulation extrapolation method for fitting generalized linear models with additive measurement error. The Stata Journal, 3(4):373–385, 2003.

Erica Ponzi, Lukas F Keller, and Stefanie Muff. The simulation extrapolation technique meets ecology and evolution: A general and intuitive method to account for measurement error. Methods in Ecology and Evolution, 10(10):1734–1748, 2019.

LA Stefanski and JM Bay. Simulation extrapolation deconvolution of finite population cumulative distribution function estimators. Biometrika, 83(2):407–417, 1996. ISSN 1464-3510.

Raymond J Carroll and Yuedong Wang. Nonparametric variance estimation in the analysis of microarray data: a measurement error approach. Biometrika, 95(2):437–449, 2008.

Xiao-Feng Wang, Zhaozhi Fan, and Bin Wang. Estimating smooth distribution function in the presence of heteroscedastic measurement errors. Computational statistics & data analysis, 54(1):25–36, 2010.

John Staudenmayer and David Ruppert. Local polynomial regression and simulation–extrapolation. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 66(1):17–30, 2004.

X.F. Wang, J.Y. Sun, and Z.Z. Fan. Deconvolution density estimation with heteroscedastic errors using simex. 2009. URL https://arxiv.org/abs/0902.2117.

R. I. Jenrich. Asymptotic properties of non-linear least squares estimators. The Annals of Mathematical Statistics, 40:633–643, 1969.

Appendix A

Appendix for Chapter 2

Proof of Theorem 1: Under (C1) through (C3), from Lemma 2 of Jenrich (1969), we know that for every $\lambda \geq 0$, there exists a measurable function $\hat{\theta}_n$ such that $\hat{\theta}_n(\lambda) = \text{argmin}_{\theta\in\Theta} L_n(\theta;\lambda)$. Moreover, based on Theorem 16(a) in Ferguson (1996), we can also show that, $\sup_{\theta\in\Theta} |L_n(\theta;\lambda) - L(\theta;\lambda)| \to 0$ with probability 1. Finally, by (C4), we can show that $\hat{\theta}_n(\lambda) \to \theta(\lambda)$ in probability. To see the approximation of $\theta(\lambda)$ to θ_0 as $\lambda \to -1$, we note that

$$
\begin{aligned}
L(\theta;\lambda) &= E\int [Y - m(Z+u;\theta)]^2 \phi(u,0,\lambda\Sigma_u) \\
&= E\int \left\{\varepsilon^2 + [m(X;\theta_0) - m(X+u;\theta)]^2\right\} \phi(u,0,(\lambda+1)\Sigma_u)du \\
&= E\int \left\{\varepsilon^2 + [m(X;\theta_0) - m(X+\sqrt{(\lambda+1)}\Sigma_u^{1/2}u;\theta)]^2\right\} \phi(u,0,I)du,
\end{aligned}
$$

it is easy to see that as $\lambda \to -1$,

$$
\sup_{\theta\in\Theta} \left| L(\theta;\lambda) - \sigma_\varepsilon^2 - \int [m(x;\theta) - m(x;\theta_0)]^2 f_X(x)dx \right| \to 0.
$$

Then from **(C4)**, we have $\theta(\lambda) \to \theta_0$ as $\lambda \to -1$.

Therefore, the derivative $\dot{L}(\theta;\lambda)$ equals

$$\iint [m(x+\sqrt{(\lambda+1)}\Sigma_u^{1/2}u;\theta) - m(x;\theta_0)]\dot{m}(x+\sqrt{(\lambda+1)}\Sigma_u^{1/2}u;\theta)\phi(u,0,I)f_X(x)dudx.$$

By Taylor expansion, we have

$$m(x+\sqrt{(\lambda+1)}\Sigma_u^{1/2}u;\theta) = m(x;\theta)+\sqrt{(\lambda+1)}u^T\Sigma_u^{1/2}m'(x;\theta)+\frac{1}{2}(\lambda+1)u'\Sigma_u^{1/2}m''(\tilde{x};\theta)\Sigma_u^{1/2}u,$$

and

$$\begin{aligned}\dot{m}(x+\sqrt{(\lambda+1)}\Sigma_u^{1/2}u;\theta) &= \dot{m}(x;\theta)+\sqrt{(\lambda+1)}\dot{m}'(x;\theta)\Sigma_u^{1/2}u \\ &\quad +\frac{1}{2}(\lambda+1)(I_{q\times q}\otimes u'\Sigma_u^{1/2})\text{diag}\left(\dot{m}_j''(x^*;\theta)\right)(I_{q\times q}\otimes\Sigma_u^{1/2}u),\end{aligned}$$

where $\dot{m}_j''(x;\theta) = \partial m''(x;\theta)/\partial\theta_j$, $j = 1,2,\ldots,q$. Therefore,

$$\begin{aligned}0 &= \dot{L}(\theta(\lambda);\lambda) \\ &= \int [m(x;\theta(\lambda)) - m(x;\theta_0)]\dot{m}(x;\theta(\lambda))f_X(x)dx \\ &\quad +\frac{1}{2}(\lambda+1)\iint [m(x;\theta(\lambda)) - m(x;\theta_0)](I_{q\times q}\otimes u'\Sigma_u^{1/2})\text{diag}\left(\dot{m}_j''(x^*;\theta)\right) \\ &\quad\quad (I_{q\times q}\otimes\Sigma_u^{1/2}u)f_X(x)\phi(u,0,I)dxdu \\ &\quad +\frac{1}{2}(\lambda+1)\iint \dot{m}(x;\theta(\lambda))u^T\Sigma_u^{1/2}m''(\tilde{x};\theta(\lambda))\Sigma_u^{1/2}uf_X(x)\phi(u,0,I)dxdu \\ &\quad +(\lambda+1)\iint u^T\Sigma_u^{1/2}m'(x;\theta(\lambda))\dot{m}'(\tilde{x};\theta(\lambda))\Sigma_u^{1/2}uf_X(x)\phi(u,0,I)dxdu \\ &\quad +o((\lambda+1)).\end{aligned}$$

Also we have $m(x;\theta(\lambda)) - m(x;\theta_0) = \dot{m}^T(x;\tilde{\theta})(\theta(\lambda)-\theta_0)$, this implies that,

$$\begin{aligned}&\left[\int \dot{m}(x;\theta_0)\dot{m}^T(x;\theta_0)f_X(x)dx + o(1)\right](\theta(\lambda)-\theta_0) \\ &= -\frac{1}{2}(\lambda+1)\iint \dot{m}(x;\theta_0)u^T\Sigma_u^{1/2}m''(x;\theta_0)\Sigma_u^{1/2}uf_X(x)\phi(u,0,I)dxdu\end{aligned}$$

$$
-(\lambda+1)\iint u^T\Sigma_u^{1/2}m'(x;\theta_0)\dot{m}'(x;\theta_0)\Sigma_u^{1/2}uf_X(x)\phi(u,0,I)dxdu+o((\lambda+1))
$$
$$
=-\frac{1}{2}(\lambda+1)\int\dot{m}(x;\theta_0)\text{trace}(m''(x;\theta_0)\Sigma_u^2)f_X(x)dx
$$
$$
-(\lambda+1)\int\dot{m}'(x;\theta_0)\Sigma_u^2m'(x;\theta_0)f_X(x)dx+o((\lambda+1)).
$$

Note that $\tilde{x}\to x$ as $\lambda\to-1$, then we obtain the approximate expansion of $\theta(\lambda)$ as $\lambda\to-1$. This concludes the proof of Theorem 1. □

Proof of Theorem 2: Denote $Q(y,z;\theta)=\int[y-m(z+u;\theta)]^2\phi(u,0,\lambda\Sigma_u)du$. Note that $\hat{\theta}_n(\lambda)$ satisfies $n^{-1}\sum_{i=1}^n\dot{Q}(Y_i,Z_i;\hat{\theta}_n(\lambda))=0$. Taylor expansion of the left hand side at $\theta=\theta(\lambda)$, the solution of $\dot{E}Q(Y,Z;\theta)=0$ leads to

$$
0=\frac{1}{n}\sum_{i=1}^n\dot{Q}(Y_i,Z_i;\theta(\lambda))+\frac{1}{n}\sum_{i=1}^n\ddot{Q}(Y_i,Z_i;\theta_n^*(\lambda))\left[\hat{\theta}_n(\lambda)-\theta(\lambda)\right],
$$

where $\theta_n^*(\lambda)$ is some value between $\hat{\theta}_n(\lambda)$ and $\theta(\lambda)$. For each $\lambda>0$, note that $\hat{\theta}_n(\lambda)\to\theta(\lambda)$ in probability implies that $\theta_n^*(\lambda)\to\theta(\lambda)$ in probability as $n\to\infty$, so

$$
\frac{1}{n}\sum_{i=1}^n\ddot{Q}(Y_i,Z_i;\theta_n^*(\lambda))\to\Sigma_1(\lambda)
$$

in probability as $n\to\infty$, where

$$
\Sigma_1(\lambda)=E\int\dot{m}(Z+u;\theta(\lambda))\dot{m}^T(Z+u;\theta(\lambda))\phi(u,0,\lambda\Sigma_u)du
$$
$$
-E\int[Y-m(Z+u,\theta(\lambda))]\ddot{m}(Z+u;\theta(\lambda))\phi(u,0,\lambda\Sigma_u)du.
$$

The asymptotic joint normality of $\hat{\theta}(\Lambda)$ is an application of the multivariate central limit theorem on $n^{-1/2}\sum_{i=1}^n\dot{Q}(Y_i,Z_i;\theta(\lambda_j)),j=1,2,\ldots,K$. It is sufficient to check the covariance matrix $\Sigma_0(\lambda_j,\lambda_l)$ of $\dot{Q}(Y,Z;\theta(\lambda_j))$ and $\dot{Q}(Y,Z;\theta(\lambda_l))$ has the specified form for $j,l=1,2,\ldots,K$. To see this, we need a general result for multivariate normal density functions. Without loss of generality, let $j=1,l=2$. Denote $\phi(x;\mu_1,\Sigma_1)$ and $\phi(x;\mu_2,\Sigma_2)$

as two p-dimensional normal density functions. Then

$$\phi(x;\mu_1,\Sigma_1)\phi(x;\mu_2,\Sigma_2) = \phi(\mu_1;\mu_2,\Sigma_1+\Sigma_2)\phi(x;\mu,\Sigma), \tag{A.0.1}$$

where $\Sigma = (\Sigma_1^{-1}+\Sigma_2^{-1})^{-1}, \mu = \Sigma(\Sigma_1^{-1}\mu_1+\Sigma_2^{-1}\mu_2)$. Thus the covariance matrix of $\dot{Q}(Y,Z;\theta(\lambda_1))$ and $\dot{Q}(Y,Z;\theta(\lambda_2))$ is

$$E\iint [Y-m(Z+u,\theta(\lambda_1))][Y-m(Z+v,\theta(\lambda_2))]\dot{m}(Z+u;\theta(\lambda_1))\dot{m}^T(Z+v;\theta(\lambda_2))$$
$$\phi(u,0,\lambda_1\Sigma_u)\phi(v,0,\lambda_2\Sigma_u)dudv.$$

By changing variables, the above expectation equals

$$\begin{aligned}
&E\iiint [Y-m(X+w+u,\theta(\lambda_1))][Y-m(X+w+v,\theta(\lambda_2))]\dot{m}(X+w+u;\theta(\lambda_1))\\
&\qquad \dot{m}^T(X+w+v;\theta(\lambda_2))\phi(u,0,\lambda_1\Sigma_u)\phi(v,0,\lambda_2\Sigma_u)\phi(w,0,\Sigma_u)dwdudv\\
=&E\iiint [Y-m(X+u,\theta(\lambda_1))][Y-m(X+v,\theta(\lambda_2))]\dot{m}(X+u;\theta(\lambda_1))\\
&\qquad \dot{m}^T(X+v;\theta(\lambda_2))\phi(w,u,\lambda_1\Sigma_u)\phi(w,v,\lambda_2\Sigma_u)\phi(w,0,\Sigma_u)dwdudv.
\end{aligned}$$

From (A.0.1), we can write $\phi(w,u,\lambda_1\Sigma_u)\phi(w,v,\lambda_2\Sigma_u)\phi(w,0,\Sigma_u)$ as either

$$\phi(u,v,(\lambda_1+\lambda_2)\Sigma_u)\phi\left(\frac{\lambda_1 v+\lambda_2 u}{\lambda_1+\lambda_2},0,\frac{\lambda_{12}}{\lambda_1+\lambda_2}\Sigma_u\right)\phi\left(w,\frac{\lambda_1 v+\lambda_2 u}{\lambda_{12}},\frac{\lambda_1\lambda_2}{\lambda_{12}}\Sigma_u\right)$$

or

$$\phi(v,0,(\lambda_2+1)\Sigma_u)\phi\left(u,\frac{v}{\lambda_2+1},\frac{\lambda_{12}}{\lambda_2+1}\Sigma_u\right)\phi\left(w,\frac{\lambda_1 v+\lambda_2 u}{\lambda_{12}},\frac{\lambda_1\lambda_2}{\lambda_{12}}\Sigma_u\right),$$

where $\lambda_{12} = \lambda_1+\lambda_2+\lambda_1\lambda_2$. This, together with $Y = m(X;\theta_0)+\varepsilon$, implies that the covariance matrix of $\dot{Q}(Y,Z;\theta(\lambda_1))$ and $\dot{Q}(Y,Z;\theta(\lambda_2))$ can be written as $\Sigma_0(\lambda_1,\lambda_2)$ as defined in Theorem 2. □

Proof of Theorem 3: Note that $\hat{\Gamma}$ is the minimizer of (2.3.4), by a Taylor expansion, we have

$$
\begin{aligned}
0 &= \dot{G}^T(\Lambda, \hat{\Gamma})(\hat{\theta}_n(\Lambda) - G(\Lambda, \hat{\Gamma})) \\
&= \dot{G}^T(\Lambda, \Gamma_0)(\hat{\theta}_n(\Lambda) - G(\Lambda, \Gamma_0)) + \left[T(\Lambda, \tilde{\Gamma}) - \dot{G}^T(\Lambda, \tilde{\Gamma})\dot{G}(\Lambda, \tilde{\Gamma})\right](\hat{\Gamma} - \Gamma_0),
\end{aligned}
$$

where $\tilde{\Gamma}$ is between $\hat{\Gamma}$ and Γ_0 and

$$
T(\Lambda, \Gamma) = \sum_{j=1}^{M} \sum_{k=1}^{q} \begin{pmatrix} \frac{\partial G_k(\lambda_j, \Gamma)}{\partial \gamma_1 \partial \Gamma^T}(\hat{\theta}_{nk}(\lambda_j) - G_k(\lambda_j, \Gamma)) \\ \vdots \\ \frac{\partial G_k(\lambda_j, \Gamma)}{\partial \gamma_d \partial \Gamma^T}(\hat{\theta}_{nk}(\lambda_j) - G_k(\lambda_j, \Gamma)) \end{pmatrix}_{d \times d},
$$

where $\hat{\theta}_{nk}(\lambda)$ is the j-th component of $\hat{\theta}_n(\lambda)$, $j = 1, 2, \ldots, q$. The consistency of $\hat{\Gamma}$ to Γ_0 implies that

$$
\sqrt{n}\left[\dot{G}^T(\Lambda, \Gamma_0)\dot{G}(\Lambda, \Gamma_0)\right](\hat{\Gamma} - \Gamma_0) = \sqrt{n}\dot{G}^T(\Lambda, \Gamma_0)(\hat{\theta}(\Lambda) - \theta(\Lambda)) + o_p(1).
$$

Recall the notations $H(\Lambda)$ and $\Pi(\Lambda)$ defined right before Theorem 3 in Section 2.3, the asymptotic normality of $\hat{\theta}_n(\Lambda)$ implies that

$$
\sqrt{n}\left[\hat{\Gamma} - \Gamma_0\right] \Longrightarrow N(0, \Pi(\Lambda)). \tag{A.0.2}
$$

Note that the EX estimate $\hat{\theta}_n$ is defined as $\hat{\theta}_n = G(-1, \hat{\Gamma})$, also note that $G(-1, \Gamma_0) = \theta_0$, so by Taylor expansion again, $\hat{\theta}_n - \theta_0 = \dot{G}(-1, \tilde{\Gamma})(\hat{\Gamma} - \Gamma_0)$, together with the asymptotic result (A.0.2), we prove Theorem 3. □

Proof of Example 2: Consider $m(x;\theta) = \exp(\theta x)$. Suppose $m(x, \theta) = \exp(\theta x)$, then we have $\dot{m}(\theta x) = x\exp(\theta x)$. Therefore, from either one of the above expression for $\Sigma_0(\lambda)$, we have

$$
\begin{aligned}
\Sigma_0(\lambda) = \sigma_\varepsilon^2 + E \iint &[\exp\theta_0 X - \exp\theta(\lambda)(X+u)] \cdot [\exp\theta_0 X - \exp\theta(\lambda)(X+v)] \cdot \\
&[(X+u)\exp(X+u)\theta(\lambda)] \cdot [(X+v)\exp\theta(\lambda)(X+v)] \cdot
\end{aligned}
$$

$$\phi(v,0,(\lambda+1)\sigma_u^2)\phi\left(u,\frac{v}{\lambda+1},\frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\right)dudv$$

$$= \sigma_\varepsilon^2 + E\iint[\exp 2\theta_0 X - \exp[X(\theta_0+\theta(\lambda))+\theta(\lambda)v]$$
$$- \exp[X(\theta_0+\theta(\lambda))+u\theta(\lambda)] + \exp[2X\theta(\lambda)+(u+v)\theta(\lambda)]]\cdot$$
$$[X^2\exp[2X\theta(\lambda)+(v+u)\theta(\lambda)] + Xv\exp[2X\theta(\lambda)+(v+u)\theta(\lambda)]$$
$$+Xu\exp[2X\theta(\lambda)+(v+u)\theta(\lambda)] + uv\exp[2X\theta(\lambda)+(v+u)\theta(\lambda)]\cdot$$
$$\phi(v,0,(\lambda+1)\sigma_u^2)\phi\left(u,\frac{v}{\lambda+1},\frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\right)dudv$$

$$= \sigma_\varepsilon^2 + E\iint X^2[\exp\left(2X(\theta_0+\theta(\lambda))+\theta(\lambda)(u+v)\right) - \exp\left(X(\theta_0+3\theta(\lambda))+(u+2v)\theta(\lambda)\right)$$
$$- \exp\left(X(\theta_0+3\theta(\lambda))+(2u+v)\theta(\lambda)\right) + \exp\left(4X\theta(\lambda)+(2u+2v)\theta(\lambda)\right)]\cdot$$
$$\phi(v,0,(\lambda+1)\sigma_u^2)\phi\left(u,\frac{v}{\lambda+1},\frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\right)dudv$$
$$+E\iint Xv[\exp\left(2X(\theta_0+\theta(\lambda))+\theta(\lambda)(u+v)\right) - \exp\left(X(\theta_0+3\theta(\lambda))+(u+2v)\theta(\lambda)\right)$$
$$- \exp\left(X(\theta_0+3\theta(\lambda))+(2u+v)\theta(\lambda)\right) + \exp\left(4X\theta(\lambda)+(2u+2v)\theta(\lambda)\right)]\cdot$$
$$\phi(v,0,(\lambda+1)\sigma_u^2)\phi\left(u,\frac{v}{\lambda+1},\frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\right)dudv$$
$$+E\iint Xu[\exp\left(2X(\theta_0+\theta(\lambda))+\theta(\lambda)(u+v)\right) - \exp\left(X(\theta_0+3\theta(\lambda))+(u+2v)\theta(\lambda)\right)$$
$$- \exp\left(X(\theta_0+3\theta(\lambda))+(2u+v)\theta(\lambda)\right) + \exp\left(4X\theta(\lambda)+(2u+2v)\theta(\lambda)\right)]\cdot$$
$$\phi(v,0,(\lambda+1)\sigma_u^2)\phi\left(u,\frac{v}{\lambda+1},\frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\right)dudv$$
$$+E\iint uv[\exp\left(2X(\theta_0+\theta(\lambda))+\theta(\lambda)(u+v)\right) - \exp\left(X(\theta_0+3\theta(\lambda))+(u+2v)\theta(\lambda)\right)$$
$$- \exp\left(X(\theta_0+3\theta(\lambda))+(2u+v)\theta(\lambda)\right) + \exp\left(4X\theta(\lambda)+(2u+2v)\theta(\lambda)\right)]\cdot$$
$$\phi(v,0,(\lambda+1)\sigma_u^2)\phi\left(u,\frac{v}{\lambda+1},\frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\right)dudv$$

Looking at the X^2-term, we have

$$\iint X^2\exp(u\theta(\lambda))\left(\exp\left(2X(\theta_0+\theta(\lambda))+v\theta(\lambda)\right) - \exp\left(X(\theta_0+3\theta(\lambda))+2v\theta(\lambda)\right)\right)\cdot$$
$$\phi(v,0,(\lambda+1)\sigma_u^2)\phi\left(u,\frac{v}{\lambda+1},\frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\right)dudv$$
$$-\iint X^2\exp(2u\theta(\lambda))\left(\exp\left(X(\theta_0+3\theta(\lambda))+v\theta(\lambda)\right) - \exp\left(4X\theta(\lambda)+2v\theta(\lambda)\right)\right)\cdot$$

$$
\begin{aligned}
&\qquad \phi(v,0,(\lambda+1)\sigma_u^2)\phi\left(u,\frac{v}{\lambda+1},\frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\right)dudv \\
&=\int X^2\exp\left(\frac{\theta(\lambda)}{\lambda+1}v+\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)+2X(\theta_0+\theta(\lambda))+v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv \\
&\quad-\int X^2\exp\left(\frac{\theta(\lambda)}{\lambda+1}v+\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)+X(\theta_0+3\theta(\lambda))+2v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv \\
&\quad-\int X^2\exp\left(\frac{2\theta(\lambda)}{\lambda+1}v+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)+X(\theta_0+3\theta(\lambda))+v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv \\
&\quad+\int X^2\exp\left(\frac{2\theta(\lambda)}{\lambda+1}v+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)+4X\theta(\lambda)+2v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv \\
&=\int X^2\exp\left(\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)+2X(\theta_0+\theta(\lambda))+\left(\frac{1}{\lambda+1}+1\right)v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv \\
&\quad-\int X^2\exp\left(\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)+X(\theta_0+3\theta(\lambda))+\left(\frac{1}{\lambda+1}+2\right)v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv \\
&\quad-\int X^2\exp\left(\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)+X(\theta_0+3\theta(\lambda))+\left(\frac{2}{\lambda+1}+1\right)v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv \\
&\quad+\int X^2\exp\left(\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)+4X\theta(\lambda)+\left(\frac{1}{\lambda+1}+1\right)2v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv \\
&=\int X^2\exp\left(\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)+2X(\theta_0+\theta(\lambda))+\frac{\lambda+2}{\lambda+1}v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv \\
&\quad-\int X^2\exp\left(\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)+X(\theta_0+3\theta(\lambda))+\frac{2\lambda+3}{\lambda+1}v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv \\
&\quad-\int X^2\exp\left(\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)+X(\theta_0+3\theta(\lambda))+\frac{\lambda+3}{\lambda+1}v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv \\
&\quad+\int X^2\exp\left(\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)+4X\theta(\lambda)+\frac{2(\lambda+2)}{\lambda+1}v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv \\
&=X^2\exp\left(\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)+2X(\theta_0+\theta(\lambda))+\frac{(\lambda+1)(\lambda+2)^2}{2(\lambda+1)^2}\sigma_u^2\theta^2(\lambda)\right) \\
&\quad-X^2\exp\left(\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)+X(\theta_0+3\theta(\lambda))+\frac{1}{2}\frac{(2\lambda+3)^2}{(\lambda+1)^2}(\lambda+1)\sigma_u^2\theta^2(\lambda)\right) \\
&\quad-X^2\exp\left(\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)+X(\theta_0+3\theta(\lambda))+\frac{1}{2}\frac{(\lambda+3)^2}{(\lambda+1)^2}(\lambda+1)\sigma_u^2\theta^2(\lambda)\right) \\
&\quad+X^2\exp\left(\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)+4X\theta(\lambda)+\frac{1}{2}\frac{(\lambda+2)^2}{(\lambda+1)^2}(\lambda+1)\sigma_u^2\theta^2(\lambda)\right) \\
&=X^2\exp((\lambda+2)\sigma_u^2\theta^2(\lambda)+2X(\theta_0+\theta(\lambda)) \\
&\quad-X^2\exp\left(\frac{\lambda^2+2\lambda+4\lambda^2+12\lambda+9)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)+X(\theta_0+3\theta(\lambda))\right) \\
&\quad-X^2\exp\left(\frac{4\lambda^2+8\lambda+\lambda^2+6\lambda+9)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)+X(\theta_0+3\theta(\lambda))\right)
\end{aligned}
$$

$$
\begin{aligned}
&+X^2\exp\left(\frac{2(\lambda+2)(\lambda+1)}{\lambda+1}\sigma_u^2\theta^2(\lambda)+4X\theta(\lambda)\right)\\
=\;& X^2\exp[2X(\theta_0+\theta(\lambda))+(\lambda+2)\sigma_u^2\theta^2(\lambda)]\\
&-X^2\exp[X(\theta_0+3\theta(\lambda))+\frac{1}{2}(5\lambda+9)\sigma_u^2\theta^2(\lambda)]\\
&+X^2\exp[4X\theta(\lambda)+4(\lambda+2)\sigma_u^2\theta^2(\lambda)]
\end{aligned}
$$

When $\lambda\to-1$, we get

$$
X^2[\exp(4X\theta_0+\sigma_u^2\theta_0^2)-2\exp(4X\theta_0+2\sigma_u^2\theta_0^2)+\exp(4X\theta_0+4\sigma_u^2\theta_0^2)]
$$

Next, we look at the $Xv-$term

$$
\begin{aligned}
&\iint Xv\exp(u\theta(\lambda))\left(\exp\left(2X(\theta_0+\theta(\lambda))+v\theta(\lambda)\right)-\exp\left(X(\theta_0+3\theta(\lambda))+2v\theta(\lambda)\right)\right)\cdot\\
&\qquad\qquad\phi(v,0,(\lambda+1)\sigma_u^2)\phi\left(u,\frac{v}{\lambda+1},\frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\right)dudv\\
&-\iint Xv\exp(2u\theta(\lambda))\left(\exp\left(X(\theta_0+3\theta(\lambda))+v\theta(\lambda)\right)-\exp\left(4X\theta(\lambda)+2v\theta(\lambda)\right)\right)\cdot\\
&\qquad\qquad\phi(v,0,(\lambda+1)\sigma_u^2)\phi\left(u,\frac{v}{\lambda+1},\frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\right)dudv\\
=\;&\int Xv\exp\left(\frac{\theta(\lambda)}{\lambda+1}v+\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)+2X(\theta_0+\theta(\lambda))+v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
&-\int Xv\exp\left(\frac{\theta(\lambda)}{\lambda+1}v+\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)+X(\theta_0+3\theta(\lambda))+2v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
&-\int Xv\exp\left(\frac{2\theta(\lambda)}{\lambda+1}v+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)+X(\theta_0+3\theta(\lambda))+v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
&+\int Xv\exp\left(\frac{2\theta(\lambda)}{\lambda+1}v+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)+4X\theta(\lambda)+2v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
=\;&\int Xv\exp\left(\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)+2X(\theta_0+\theta(\lambda))+\left(\frac{1}{\lambda+1}+1\right)v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
&-\int Xv\exp\left(\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)+X(\theta_0+3\theta(\lambda))+\left(\frac{1}{\lambda+1}+2\right)v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
&-\int Xv\exp\left(\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)+X(\theta_0+3\theta(\lambda))+\left(\frac{2}{\lambda+1}+1\right)v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
&+\int Xv\exp\left(\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)+4X\theta(\lambda)+\left(\frac{1}{\lambda+1}+1\right)2v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv
\end{aligned}
$$

$$
\begin{aligned}
&= \int Xv \exp\left(\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda) + 2X(\theta_0+\theta(\lambda)) + \frac{\lambda+2}{\lambda+1}v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv \\
&\quad - \int Xv \exp\left(\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda) + X(\theta_0+3\theta(\lambda)) + \frac{2\lambda+3}{\lambda+1}v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv \\
&\quad - \int Xv \exp\left(\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda) + X(\theta_0+3\theta(\lambda)) + \frac{\lambda+3}{\lambda+1}v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv \\
&\quad + \int Xv \exp\left(\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda) + 4X\theta(\lambda) + \frac{2(\lambda+2)}{\lambda+1}v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv \\
&= \frac{\lambda+1}{\lambda+2}X\exp\left(\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda) + 2X(\theta_0+\theta(\lambda))\right)\frac{\partial}{\partial\theta(\lambda)}\int \exp\left(\frac{\lambda+2}{\lambda+1}v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv \\
&\quad - \frac{\lambda+1}{2\lambda+3}X\exp\left(\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda) + X(\theta_0+3\theta(\lambda))\right)\frac{\partial}{\partial\theta(\lambda)}\int \exp\left(\frac{2\lambda+3}{\lambda+1}v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv \\
&\quad - \frac{\lambda+1}{\lambda+3}X\exp\left(\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda) + X(\theta_0+3\theta(\lambda))\right)\frac{\partial}{\partial\theta(\lambda)}\int \left(\frac{\lambda+3}{\lambda+1}v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv \\
&\quad + \frac{\lambda+1}{2(\lambda+2)}X\exp\left(\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda) + 4X\theta(\lambda)\right)\frac{\partial}{\partial\theta(\lambda)}\int \exp\left(\frac{2(\lambda+2)}{\lambda+1}v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv \\
&= \frac{\lambda+1}{\lambda+2}X\exp\left(\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda) + 2X(\theta_0+\theta(\lambda))\right)\cdot\frac{\partial}{\partial\theta(\lambda)}\exp\left(\frac{(\lambda+1)(\lambda+2)^2}{2(\lambda+1)^2}\sigma_u^2\theta^2(\lambda)\right) \\
&\quad - \frac{\lambda+1}{2\lambda+3}X\exp\left(\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda) + X(\theta_0+3\theta(\lambda))\right)\cdot\frac{\partial}{\partial\theta(\lambda)}\exp\left(\frac{1}{2}\frac{(2\lambda+3)^2}{(\lambda+1)^2}(\lambda+1)\sigma_u^2\theta^2(\lambda)\right) \\
&\quad - \frac{\lambda+1}{\lambda+3}X\exp\left(\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda) + X(\theta_0+3\theta(\lambda))\right)\cdot\frac{\partial}{\partial\theta(\lambda)}\exp\left(\frac{1}{2}\frac{(\lambda+3)^2}{(\lambda+1)^2}(\lambda+1)\sigma_u^2\theta^2(\lambda)\right) \\
&\quad + \frac{\lambda+1}{2(\lambda+2)}X\exp\left(\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda) + 4X\theta(\lambda)\right)\cdot\frac{\partial}{\partial\theta(\lambda)}\exp\left(2\frac{(\lambda+2)^2}{(\lambda+1)^2}(\lambda+1)\sigma_u^2\theta^2(\lambda)\right) \\
&= \frac{\lambda+1}{\lambda+2}X\exp\left(\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda) + 2X(\theta_0+\theta(\lambda))\right)\cdot\frac{\partial}{\partial\theta(\lambda)}\exp\left(\frac{(\lambda+2)^2}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)\right) \\
&\quad - \frac{\lambda+1}{2\lambda+3}X\exp\left(\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda) + X(\theta_0+3\theta(\lambda))\right)\cdot\frac{\partial}{\partial\theta(\lambda)}\exp\left(\frac{1}{2}\frac{(2\lambda+3)^2}{\lambda+1}\sigma_u^2\theta^2(\lambda)\right) \\
&\quad - \frac{\lambda+1}{\lambda+3}X\exp\left(\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda) + X(\theta_0+3\theta(\lambda))\right)\cdot\frac{\partial}{\partial\theta(\lambda)}\exp\left(\frac{1}{2}\frac{(\lambda+3)^2}{\lambda+1}\sigma_u^2\theta^2(\lambda)\right) \\
&\quad + 2\frac{\lambda+1}{2(\lambda+2)}X\exp\left(\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda) + 4X\theta(\lambda)\right)\cdot\frac{\partial}{\partial\theta(\lambda)}\exp\left(2\frac{(\lambda+2)^2}{\lambda+1}\sigma_u^2\theta^2(\lambda)\right) \\
&= \frac{\lambda+1}{\lambda+2}\cdot\frac{(\lambda+2)^2}{\lambda+1}\sigma_u^2\theta(\lambda)X\exp\left(\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda) + 2X(\theta_0+\theta(\lambda)) + \frac{(\lambda+2)^2}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)\right) \\
&\quad - \frac{\lambda+1}{2\lambda+3}\cdot\frac{(2\lambda+3)^2}{\lambda+1}\sigma_u^2\theta(\lambda)X\exp\left(\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda) + X(\theta_0+3\theta(\lambda)) + \frac{1}{2}\frac{(2\lambda+3)^2}{\lambda+1}\sigma_u^2\theta^2(\lambda)\right) \\
&\quad - \frac{\lambda+1}{\lambda+3}\cdot\frac{(\lambda+3)^2}{\lambda+1}\sigma_u^2\theta(\lambda)X\exp\left(\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda) + X(\theta_0+3\theta(\lambda)) + \frac{1}{2}\frac{(\lambda+3)^2}{\lambda+1}\sigma_u^2\theta^2(\lambda)\right) \\
&\quad + 2\frac{\lambda+1}{(\lambda+2)}\cdot\frac{(\lambda+2)^2}{\lambda+1}\sigma_u^2\theta(\lambda)X\exp\left(\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda) + 4X\theta(\lambda) + 2\frac{(\lambda+2)^2}{\lambda+1}\sigma_u^2\theta^2(\lambda)\right)
\end{aligned}
$$

$$= (\lambda+2)\sigma_u^2\theta(\lambda)X \exp\left(2X(\theta_0+\theta(\lambda)) + (\lambda+2)\sigma_u^2\theta^2(\lambda)\right)$$
$$-(2\lambda+3)\sigma_u^2\theta(\lambda)X \exp\left(X(\theta_0+3\theta(\lambda)) + \frac{1}{2}(5\lambda+9)\sigma_u^2\theta^2(\lambda)\right)$$
$$-(\lambda+3)\sigma_u^2\theta(\lambda)X \exp\left(X(\theta_0+3\theta(\lambda)) + \frac{1}{2}(5\lambda+3)^9\sigma_u^2\theta^2(\lambda)\right)$$
$$+2(\lambda+2)\sigma_u^2\theta(\lambda)X \exp\left(4X\theta(\lambda) + 4(\lambda+2)\sigma_u^2\theta^2(\lambda)\right)$$
$$= (\lambda+2)\sigma_u^2\theta(\lambda)X \exp\left(2X(\theta_0+\theta(\lambda)) + (\lambda+2)\sigma_u^2\theta^2(\lambda)\right)$$
$$-3(\lambda+2)\sigma_u^2\theta(\lambda)X \exp\left(X(\theta_0+3\theta(\lambda)) + \frac{1}{2}(5\lambda+9)\sigma_u^2\theta^2(\lambda)\right)$$
$$+2(\lambda+2)\sigma_u^2\theta(\lambda)X \exp\left(4X\theta(\lambda) + 4(\lambda+2)\sigma_u^2\theta^2(\lambda)\right)$$

When $\lambda \to -1$, we get

$$\sigma_u^2\theta_0 X \exp\left(4X\theta_0 + \sigma_u^2\theta_0^2\right) - 3\sigma_u^2\theta_0 X \exp\left(4X\theta_0 + 2\sigma_u^2\theta_0^2\right) + 2\sigma_u^2\theta_0 X \exp\left(4X\theta_0 + 4\sigma_u^2\theta_0^2\right)$$

Next, we look at the $Xu-$term, i.e.,

$$\iint Xu \exp(u\theta(\lambda)) \left(\exp\left(2X(\theta_0+\theta(\lambda)) + v\theta(\lambda)\right) - \exp\left(X(\theta_0+3\theta(\lambda)) + 2v\theta(\lambda)\right)\right)$$
$$\cdot\phi(v,0,(\lambda+1)\sigma_u^2)\phi\left(u, \frac{v}{\lambda+1}, \frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\right) dudv$$
$$-\iint Xu \exp(2u\theta(\lambda)) \left(\exp\left(X(\theta_0+3\theta(\lambda)) + v\theta(\lambda)\right) - \exp\left(4X\theta(\lambda) + 2v\theta(\lambda)\right)\right)$$
$$\cdot\phi(v,0,(\lambda+1)\sigma_u^2)\phi\left(u, \frac{v}{\lambda+1}, \frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\right) dudv$$
$$= \int X \left(\exp\left(2X(\theta_0+\theta(\lambda)) + v\theta(\lambda)\right) - \exp\left(X(\theta_0+3\theta(\lambda)) + 2v\theta(\lambda)\right)\right)$$
$$\cdot\left[\frac{\partial}{\theta(\lambda)} \int \exp(u\theta(\lambda))\phi\left(u, \frac{v}{\lambda+1}, \frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\right) du\right] \phi(v,0,(\lambda+1)\sigma_u^2)dv$$
$$-\int \frac{1}{2} X \left(\exp\left(X(\theta_0+3\theta(\lambda)) + v\theta(\lambda)\right) - \exp\left(4X\theta(\lambda) + 2v\theta(\lambda)\right)\right)$$
$$\cdot\left[\frac{\partial}{\theta(\lambda)} \int \exp(2u\theta(\lambda))\phi\left(u, \frac{v}{\lambda+1}, \frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\right) du\right] \phi(v,0,(\lambda+1)\sigma_u^2)dv$$
$$= \int X \left(\exp\left(2X(\theta_0+\theta(\lambda)) + v\theta(\lambda)\right) - \exp\left(X(\theta_0+3\theta(\lambda)) + 2v\theta(\lambda)\right)\right)$$
$$\cdot\frac{\partial}{\theta(\lambda)} \left[\exp\left(\frac{\theta(\lambda)}{\lambda+1}v + \frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)\right)\right] \phi(v,0,(\lambda+1)\sigma_u^2)dv$$

$$
\begin{aligned}
&-\int \frac{1}{2}X\left(\exp\left(X(\theta_0+3\theta(\lambda))+v\theta(\lambda)\right)-\exp\left(4X\theta(\lambda)+2v\theta(\lambda)\right)\right)\\
&\qquad\cdot\frac{\partial}{\theta(\lambda)}\left[\exp\left(\frac{2\theta(\lambda)}{\lambda+1}v+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)\right)\right]\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
=&\int X\left(\exp\left(2X(\theta_0+\theta(\lambda))+v\theta(\lambda)\right)-\exp\left(X(\theta_0+3\theta(\lambda))+2v\theta(\lambda)\right)\right)\\
&\qquad\cdot\left[\frac{v}{\lambda+1}+\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta(\lambda)\right]\left[\exp\left(\frac{\theta(\lambda)}{\lambda+1}v+\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)\right)\right]\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
&-\frac{1}{2}\int X\left(\exp\left(X(\theta_0+3\theta(\lambda))+v\theta(\lambda)\right)-\exp\left(4X\theta(\lambda)+2v\theta(\lambda)\right)\right)\\
&\qquad\cdot\left[\frac{2v}{\lambda+1}+\frac{4\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta(\lambda)\right]\left[\exp\left(\frac{2\theta(\lambda)}{\lambda+1}v+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)\right)\right]\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
=&X\exp\left(2X(\theta_0+\theta(\lambda)+\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)\right)\\
&\qquad\cdot\int\left(\frac{v}{\lambda+1}+\frac{\lambda(\lambda+2)}{(\lambda+1)}\sigma_u^2\theta(\lambda)\right)\exp\left(v\theta(\lambda))+\frac{v\theta(\lambda)}{\lambda+1}\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
&-X\exp\left(X(\theta_0+3\theta(\lambda)+\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)\right)\\
&\qquad\cdot\int\left(\frac{v}{\lambda+1}+\frac{\lambda(\lambda+2)}{(\lambda+1)}\sigma_u^2\theta(\lambda)\right)\exp\left(2v\theta(\lambda))+\frac{v\theta(\lambda)}{\lambda+1}\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
&-\frac{1}{2}X\exp\left(X(\theta_0+3\theta(\lambda)+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)\right)\\
&\qquad\cdot\int\left(\frac{2v}{\lambda+1}+\frac{4\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta(\lambda)\right)\exp\left[v\theta(\lambda)\left(1+\frac{2}{\lambda+1}\right)\right]\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
&+\frac{1}{2}X\exp\left(4X\theta(\lambda)+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)\right)\\
&\qquad\cdot\int\left(\frac{2v}{\lambda+1}+\frac{4\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta(\lambda)\right)\exp\left[2v\theta(\lambda)\left(1+\frac{1}{\lambda+1}\right)\right]\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
=&Xe^{2X(\theta_0+\theta(\lambda))+\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)}\int\frac{v}{(\lambda+1)}\exp\left(\frac{\lambda+2}{\lambda+1}v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
&+X\frac{\lambda(\lambda+2)}{(\lambda+1)}\sigma_u^2\theta(\lambda)e^{2X(\theta_0+\theta(\lambda)+\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)}\int\exp\left(\frac{\lambda+2}{\lambda+1}v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
&-Xe^{X(\theta_0+3\theta(\lambda))+\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)}\int\frac{v}{(\lambda+1)}\exp\left(\frac{2\lambda+3}{\lambda+1}v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
&-X\frac{\lambda(\lambda+2)}{(\lambda+1)}\sigma_u^2\theta(\lambda)e^{X(\theta_0+3\theta(\lambda)+\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)}\int\exp\left(\frac{2\lambda+3}{\lambda+1}v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
&-Xe^{X(\theta_0+3\theta(\lambda))+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)}\int\frac{v}{(\lambda+1)}\exp\left(\frac{\lambda+3}{\lambda+1}v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
&-X\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta(\lambda)e^{X(\theta_0+3\theta(\lambda)+\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)}\int\exp\left(\frac{\lambda+3}{\lambda+1}v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{1}{2}Xe^{4X\theta(\lambda)+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)}\int\frac{2v}{(\lambda+1)}\exp\left(\frac{\lambda+2}{\lambda+1}2v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
&+\frac{1}{2}X\frac{4\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta(\lambda)e^{4X\theta(\lambda)+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)}\int\exp\left(\frac{\lambda+2}{\lambda+1}2v\theta(\lambda)\right)\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
&=Xe^{2X(\theta_0+\theta(\lambda))+\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)}\frac{1}{\lambda+2}\left[\frac{\partial}{\partial\theta(\lambda)}\exp\left(\frac{1}{2}\frac{(\lambda+2)^2}{\lambda+1}\theta^2(\lambda)\sigma_u^2\right)\right]\\
&+\frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta(\lambda)X\exp\left(2X(\theta_0+\theta(\lambda))+\frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)+\frac{1}{2}\frac{(\lambda+2)^2}{\lambda+1}\theta^2(\lambda)\sigma_u^2\right)\\
&-Xe^{X(\theta_0+3\theta(\lambda))+\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)}\frac{1}{2\lambda+3}\left[\frac{\partial}{\partial\theta(\lambda)}\exp\left(\frac{1}{2}\frac{(2\lambda+3)^2}{\lambda+1}\theta^2(\lambda)\sigma_u^2\right)\right]\\
&-\frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta(\lambda)X\exp\left(X(\theta_0+3\theta(\lambda))+\frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)+\frac{1}{2}\frac{(2\lambda+3)^2}{\lambda+1}\theta^2(\lambda)\sigma_u^2\right)\\
&-Xe^{X(\theta_0+3\theta(\lambda))+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)}\frac{1}{\lambda+3}\left[\frac{\partial}{\partial\theta(\lambda)}\exp\left(\frac{1}{2}\frac{(\lambda+3)^2}{\lambda+1}\sigma_u^2\theta^2(\lambda)\right)\right]\\
&-\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta(\lambda)X\exp\left(X(\theta_0+3\theta(\lambda))+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)+\frac{1}{2}\frac{(\lambda+3)^2}{\lambda+1}\sigma_u^2\theta^2(\lambda)\right)\\
&+\frac{1}{2}Xe^{4X\theta(\lambda)+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)}\frac{1}{\lambda+2}\left[\frac{\partial}{\partial\theta(\lambda)}\exp\left(2\frac{(\lambda+2)^2}{\lambda+1}\sigma_u^2\theta^2(\lambda)\right)\right]\\
&+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta(\lambda)X\exp\left(4X\theta(\lambda)+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)+2\frac{(\lambda+2)^2}{\lambda+1}\sigma_u^2\theta^2(\lambda)\right)\\
&=\left[\frac{\lambda+2}{\lambda+1}+\frac{\lambda(\lambda+2)}{\lambda+1}\right]\sigma_u^2\theta(\lambda)X\exp\left(2X(\theta_0+\theta(\lambda))+\frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)+\frac{1}{2}\frac{(\lambda+2)^2}{\lambda+1}\theta^2(\lambda)\sigma_u^2\right)\\
&-\left[\frac{2\lambda+3}{\lambda+1}+\frac{\lambda(\lambda+2)}{\lambda+1}\right]\sigma_u^2\theta(\lambda)X\exp\left(X(\theta_0+3\theta(\lambda))+\frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)+\frac{1}{2}\frac{(2\lambda+3)^2}{\lambda+1}\theta^2(\lambda)\sigma_u^2\right)\\
&-\left[\frac{\lambda+3}{\lambda+1}+\frac{2\lambda(\lambda+2)}{\lambda+1}\right]\sigma_u^2\theta(\lambda)X\exp\left(X(\theta_0+3\theta(\lambda))+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)+\frac{1}{2}\frac{(\lambda+3)^2}{\lambda+1}\sigma_u^2\theta^2(\lambda)\right)\\
&+\left[\frac{\lambda+2}{\lambda+1}+\frac{\lambda(\lambda+2)}{\lambda+1}\right]\sigma_u^2\theta(\lambda)X\exp\left(4X\theta(\lambda)+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)+2\frac{(\lambda+2)^2}{\lambda+1}\sigma_u^2\theta^2(\lambda)\right)\\
&=(\lambda+2)\sigma_u^2\theta(\lambda)X\exp[2X(\theta_0+\theta(\lambda))+(\lambda+2)\sigma_u^2\theta^2(\lambda)]\\
&-(\lambda+3)\sigma_u^2\theta(\lambda)X\exp[X(\theta_0+3\theta(\lambda))+\frac{1}{2}(5\lambda+9)\sigma_u^2\theta^2(\lambda)]\\
&-(2\lambda+3)\sigma_u^2\theta(\lambda)X\exp[X(\theta_0+3\theta(\lambda))+\frac{1}{2}(5\lambda+9)\sigma_u^2\theta^2(\lambda)]\\
&+2(\lambda+2)\sigma_u^2\theta(\lambda)X\exp[4X\theta(\lambda)+4(\lambda+2)\sigma_u^2\theta^2(\lambda)]
\end{aligned}
$$

When $\lambda\to-1$, we get

$$
\sigma_u^2\theta_0X\exp\left(4X\theta_0+\sigma_u^2\theta_0^2\right)-3\sigma_u^2\theta_0X\exp\left(4X\theta_0+2\sigma_u^2\theta_0^2\right)+2\sigma_u^2\theta_0X\exp\left(4X\theta_0+4\sigma_u^2\theta_0^2\right)
$$

Looking at the $uv-$term, we get

$$
\begin{aligned}
&\int v\left(\exp\left(2X(\theta_0+\theta(\lambda))+v\theta(\lambda)\right)-\exp\left(X(\theta_0+3\theta(\lambda))+2v\theta(\lambda)\right)\right)\\
&\qquad\cdot\left[\int u\exp(u\theta(\lambda))\phi\left(u,\frac{v}{\lambda+1},\frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\right)du\right]\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
&-\int v\left(\exp\left(X(\theta_0+3\theta(\lambda))+v\theta(\lambda)\right)-\exp\left(4X\theta(\lambda)+2v\theta(\lambda)\right)\right)\\
&\qquad\cdot\left[\int u\exp(2u\theta(\lambda))\phi\left(u,\frac{v}{\lambda+1},\frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\right)du\right]\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
=&\int v\left(\exp\left(2X(\theta_0+\theta(\lambda))+v\theta(\lambda)\right)-\exp\left(X(\theta_0+3\theta(\lambda))+2v\theta(\lambda)\right)\right)\\
&\qquad\cdot\left[\frac{\partial}{\partial\theta(\lambda)}\int\exp(u\theta(\lambda))\phi\left(u,\frac{v}{\lambda+1},\frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\right)du\right]\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
&-\frac{1}{2}\int v\left(\exp\left(X(\theta_0+3\theta(\lambda))+v\theta(\lambda)\right)-\exp\left(4X\theta(\lambda)+2v\theta(\lambda)\right)\right)\\
&\qquad\cdot\left[\frac{\partial}{\partial\theta(\lambda)}\int\exp(2u\theta(\lambda))\phi\left(u,\frac{v}{\lambda+1},\frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\right)du\right]\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
=&\int v\left(\exp\left(2X(\theta_0+\theta(\lambda))+v\theta(\lambda)\right)-\exp\left(X(\theta_0+3\theta(\lambda))+2v\theta(\lambda)\right)\right)\\
&\qquad\cdot\left[\frac{\partial}{\partial\theta(\lambda)}\exp\left(\frac{v}{\lambda+1}\theta(\lambda)+\frac{1}{2}\frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)\right)\right]\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
&-\frac{1}{2}\int v\left(\exp\left(X(\theta_0+3\theta(\lambda))+v\theta(\lambda)\right)-\exp\left(4X\theta(\lambda)+2v\theta(\lambda)\right)\right)\\
&\qquad\cdot\left[\frac{\partial}{\partial\theta(\lambda)}\exp\left(\frac{2v}{\lambda+1}\theta(\lambda)+\frac{1}{2}\frac{4\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)\right)\right]\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
=&\,e^{2X(\theta_0+\theta(\lambda)+\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)}\int\left(\frac{v}{\lambda+1}+\frac{\lambda(\lambda+2)}{(\lambda+1)}\sigma_u^2\theta(\lambda)\right)ve^{v\theta(\lambda))+\frac{v\theta(\lambda)}{\lambda+1}}\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
&-e^{X(\theta_0+3\theta(\lambda)+\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)}\int\left(\frac{v}{\lambda+1}+\frac{\lambda(\lambda+2)}{(\lambda+1)}\sigma_u^2\theta(\lambda)\right)ve^{2v\theta(\lambda))+\frac{v\theta(\lambda)}{\lambda+1}}\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
&-\frac{1}{2}e^{X(\theta_0+3\theta(\lambda)+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)}\int\left(\frac{2v}{\lambda+1}+\frac{4\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta(\lambda)\right)ve^{v\theta(\lambda)\left(1+\frac{2}{\lambda+1}\right)}\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
&+\frac{1}{2}e^{4X\theta(\lambda)+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)}\int\left(\frac{2v}{\lambda+1}+\frac{4\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta(\lambda)\right)ve^{2v\theta(\lambda)\left(1+\frac{1}{\lambda+1}\right)}\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
=&\,e^{2X(\theta_0+\theta(\lambda)+\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)}\int\frac{v^2}{\lambda+1}e^{\frac{\lambda+2}{\lambda+1}v\theta(\lambda))}\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
&+\frac{\lambda(\lambda+2)}{(\lambda+1)}\sigma_u^2\theta(\lambda)e^{2X(\theta_0+\theta(\lambda)+\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)}\int ve^{\frac{\lambda+2}{\lambda+1}v\theta(\lambda))}\phi(v,0,(\lambda+1)\sigma_u^2)dv\\
&-e^{X(\theta_0+3\theta(\lambda)+\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)}\int\frac{v^2}{\lambda+1}e^{\frac{2\lambda+3}{\lambda+1}v\theta(\lambda))}\phi(v,0,(\lambda+1)\sigma_u^2)dv
\end{aligned}
$$

$$
\begin{aligned}
&-\frac{\lambda(\lambda+2)}{(\lambda+1)}\sigma_u^2\theta(\lambda)e^{X(\theta_0+3\theta(\lambda)+\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)}\int ve^{\frac{2\lambda+3}{\lambda+1}v\theta(\lambda))}\phi(v,0,(\lambda+1)\sigma_u^2)dv \\
&-e^{X(\theta_0+3\theta(\lambda)+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)}\int \frac{v^2}{\lambda+1}e^{\frac{\lambda+3}{\lambda+1}v\theta(\lambda))}\phi(v,0,(\lambda+1)\sigma_u^2)dv \\
&-\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta(\lambda)e^{X(\theta_0+3\theta(\lambda)+\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)}\int ve^{\frac{\lambda+3}{\lambda+1}v\theta(\lambda))}\phi(v,0,(\lambda+1)\sigma_u^2)dv \\
&+\frac{1}{2}e^{4X\theta(\lambda)+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)}\int \frac{2v^2}{\lambda+1}e^{2v\theta(\lambda)\frac{\lambda+2}{\lambda+1}}\phi(v,0,(\lambda+1)\sigma_u^2)dv \\
&+\frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta(\lambda)e^{4X\theta(\lambda)+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)}\int 2ve^{2v\theta(\lambda)\frac{\lambda+2}{\lambda+1}}\phi(v,0,(\lambda+1)\sigma_u^2)dv \\
=&\frac{\lambda+1}{(\lambda+2)^2}e^{2X(\theta_0+\theta(\lambda))+\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)}\left[\frac{\partial^2}{\partial\theta^2(\lambda)}\exp\left(\frac{1}{2}\frac{(\lambda+2)^2}{\lambda+1}\theta^2(\lambda)\sigma_u^2\right)\right] \\
&+\lambda\sigma_u^2\theta(\lambda)e^{2X(\theta_0+\theta(\lambda))+\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)}\left[\frac{\partial}{\partial\theta(\lambda)}\exp\left(\frac{1}{2}\frac{(\lambda+2)^2}{\lambda+1}\theta^2(\lambda)\sigma_u^2\right)\right] \\
&-\frac{\lambda+1}{(2\lambda+3)^2}e^{X(\theta_0+3\theta(\lambda))+\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)}\left[\frac{\partial^2}{\partial\theta^2(\lambda)}\exp\left(\frac{1}{2}\frac{(2\lambda+3)^2}{\lambda+1}\theta^2(\lambda)\sigma_u^2\right)\right] \\
&-\frac{\lambda(\lambda+2)}{2\lambda+3}\sigma_u^2\theta(\lambda)e^{X(\theta_0+3\theta(\lambda))+\frac{1}{2}\frac{\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)}\left[\frac{\partial}{\partial\theta(\lambda)}\exp\left(\frac{1}{2}\frac{(2\lambda+3)^2}{\lambda+1}\theta^2(\lambda)\sigma_u^2\right)\right] \\
&-\frac{\lambda+1}{(\lambda+3)^2}e^{X(\theta_0+3\theta(\lambda))+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)}\left[\frac{\partial^2}{\partial\theta^2(\lambda)}\exp\left(\frac{1}{2}\frac{(\lambda+3)^2}{\lambda+1}\sigma_u^2\theta^2(\lambda)\right)\right] \\
&-\frac{2\lambda(\lambda+2)}{\lambda+3}e^{X(\theta_0+3\theta(\lambda))+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)}\left[\frac{\partial}{\partial\theta(\lambda)}\exp\left(\frac{1}{2}\frac{(\lambda+3)^2}{\lambda+1}\sigma_u^2\theta^2(\lambda)\right)\right] \\
&+\frac{1}{4}\frac{\lambda+1}{(\lambda+2)^2}e^{4X\theta(\lambda)+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)}\left[\frac{\partial^2}{\partial\theta^2(\lambda)}\exp\left(2\frac{(\lambda+2)^2}{\lambda+1}\theta^2(\lambda)\sigma_u^2\right)\right] \\
&+\lambda\sigma_u^2\theta(\lambda)e^{4X\theta(\lambda)+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)}\left[\frac{\partial}{\partial\theta(\lambda)}\exp\left(2\frac{(\lambda+2)^2}{\lambda+1}\theta^2(\lambda)\sigma_u^2\right)\right] \\
=&\left[\frac{(\lambda+2)^2}{\lambda+1}\sigma_u^4\theta^2(\lambda)+\frac{\lambda(\lambda+2)^2}{\lambda+1}\sigma_u^4\theta^2(\lambda)+\sigma_u^2\right] \\
&\qquad\cdot\exp\left(2X(\theta_0+\theta(\lambda))+\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)+\frac{1}{2}\frac{(\lambda+2)^2}{\lambda+1}\theta^2(\lambda)\sigma_u^2\right) \\
&-\left[\frac{(2\lambda+3)^2}{\lambda+1}\sigma_u^4\theta^2(\lambda)+\frac{\lambda(\lambda+2)(2\lambda+3)}{\lambda+1}\sigma_u^4\theta^2(\lambda)+\sigma_u^2\right] \\
&\qquad\cdot\exp\left(X(\theta_0+3\theta(\lambda))+\frac{\lambda(\lambda+2)}{2(\lambda+1)}\sigma_u^2\theta^2(\lambda)+\frac{1}{2}\frac{(2\lambda+3)^2}{\lambda+1}\theta^2(\lambda)\sigma_u^2\right) \\
&-\left[\frac{(\lambda+3)^2}{\lambda+1}\sigma_u^4\theta^2(\lambda)+\frac{2\lambda(\lambda+2)(\lambda+3)}{\lambda+1}\sigma_u^4\theta^2(\lambda)+\sigma_u^2\right] \\
&\qquad\cdot\exp\left(X(\theta_0+3\theta(\lambda))+\frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda)+\frac{1}{2}\frac{(\lambda+3)^2}{\lambda+1}\theta^2(\lambda)\sigma_u^2\right) \\
&+\left[\frac{4(\lambda+2)^2}{\lambda+1}\sigma_u^4\theta^2(\lambda)+\frac{4\lambda(\lambda+2)^2}{\lambda+1}\sigma_u^4\theta^2(\lambda)+\sigma_u^2\right]
\end{aligned}
$$

$$
\begin{aligned}
&\cdot \exp\left(4X\theta(\lambda) + \frac{2\lambda(\lambda+2)}{\lambda+1}\sigma_u^2\theta^2(\lambda) + 2\frac{(\lambda+2)^2}{\lambda+1}\theta^2(\lambda)\sigma_u^2\right) \\
= &\left[\sigma_u^2 + (\lambda+2)^2\sigma_u^4\theta^2(\lambda)\right]\exp\left(2X(\theta_0+\theta(\lambda)) + (\lambda+2)\sigma_u^2\theta^2(\lambda)\right) \\
&-2\left[\sigma_u^2 + (2\lambda+3)(\lambda+3)\sigma_u^4\theta^2(\lambda)\right]\exp\left(X(\theta_0+3\theta(\lambda)) + \frac{1}{2}(5\lambda+9)\sigma_u^2\theta^2(\lambda)\right) \\
&+\left[\sigma_u^2 + 4(\lambda+2)^2\sigma_u^4\theta^2(\lambda)\right]\exp\left(4X\theta(\lambda) + 4(\lambda+2)\sigma_u^2\theta^2(\lambda)\right)
\end{aligned}
$$

when $\lambda \to -1$, we get

$$
\left[\sigma_u^2 + \sigma_u^4\theta_0^2\right]e^{4X\theta_0+\sigma_u^2\theta_0^2} - 2\left[\sigma_u^2 + 2\sigma_u^4\theta_0^2\right]e^{4X\theta_0+2\sigma_u^2\theta_0^2} + \left[\sigma_u^2 + 4\sigma_u^4\theta_0^2\right]e^{4X\theta_0+4\sigma_u^2\theta_0^2}
$$

Appendix B

Appendix for Chapter 3

B.1 Appendix

To prove Theorem 4, we need find out the expectations and variances of each component appearing in $\hat{g}_n(x;\lambda)$ defined in (3.3.8). The calculation is facilitated by the following lemmas.

Lemma 7. *Let a, c be any positive constants. Suppose that for any x in the support of f_X, $m'(t+x), m''(t+x) \in L_1(\phi(\cdot;0,c))$ and are continuous as functions of t. Then, as $h \to 0$, we have*

$$\int \phi(t;x,ah^2+c)m(t)dt = \int \phi(t;x,c)m(t)dt + \frac{ah^2}{2}\int m''(t)\phi(t;x,c)dt + o(h^2).$$

Furthermore, if

$$\frac{\partial^j \int m(t+x)\phi(t;0,c)dt}{\partial x^j} = \int \frac{\partial^j m(t+x)}{\partial x^j}\phi(t;0,c)dt, \quad j=1,2,$$

then we have

$$\int \phi(t;x,ah^2+c)m(t)dt = \int \phi(t-x;0,c)m(t)dt + \frac{ah^2}{2}\cdot\frac{\partial^2 \int m(t)\phi(t-x;0,c)dt}{\partial x^2} + o(h^2).$$

In Lemma 7, if we take $c = 0$, then

$$\int \phi(t; x, ah^2)m(t)dt = m(x) + \frac{ah^2}{2} m''(x) + o(h^2).$$

The proof of Lemma 7. Note that

$$\frac{\partial m(x + u\sqrt{ah^2 + c})}{\partial h}\Big|_{h=0} = 0, \qquad \frac{\partial^2 m(x + u\sqrt{ah^2 + c})}{\partial h^2}\Big|_{h=0} = m'(x + u\sqrt{c})\frac{au}{\sqrt{c}}.$$

Therefore, using Taylor expansion,

$$\begin{aligned} &\int \phi(t; x, ah^2 + c)m(t)dt = \int \phi(u; 0, 1)m(x + u\sqrt{ah^2 + c})du \\ &= \int \phi(u; 0, 1)\left[m(x + u\sqrt{c}) + m'(x + u\sqrt{c})\frac{auh^2}{2\sqrt{c}}\right] du + o(h^2) \\ &= \int \phi(t - x; 0, c)\left[m(t) + m'(t)\frac{ah^2(t - x)}{2c}\right] dt + o(h^2). \end{aligned} \tag{B.1.1}$$

Note that under the condition of $m'(t + x) \in L_1(\phi(\cdot; 0, c))$ for any $x \in \mathbb{R}$, we can get

$$\frac{1}{c}\int \phi(t - x; 0, c)m'(t)(t - x)dt = \int m''(t)\phi(t - x; 0, c)dt.$$

This, together with (B.1.1), implies the first expansion. For the second expansion, notice that under the derivative-integration exchangeability condition, we have

$$\int m''(t)\phi(t - x; 0, c)dt = \int \frac{\partial^2 m(t + x)}{\partial t^2}\phi(t; 0, c)dt = \int \frac{\partial^2 m(t + x)}{\partial x^2}\phi(t; 0, c)dt.$$

□

The following lemma lists some facts about normal density functions which are used often in the proofs of our main results. For the sake of brevity, the proofs of these facts are omitted since they can be found in standard statistics books.

Lemma 8. *For normal density function $\phi(u;\mu,\sigma^2)$ with mean μ and variance σ^2, we have*

$$\phi^2(u;\mu,\sigma^2) = \frac{1}{2\sqrt{\pi\sigma^2}}\phi\left(u;\mu,\frac{\sigma^2}{2}\right), \quad \phi^3(u;\mu,\sigma^2) = \frac{1}{2\sqrt{3}\pi\sigma^2}\phi\left(u;\mu,\frac{\sigma^2}{3}\right),$$
$$\phi(u;\mu_1,\sigma_1^2)\phi(u,\mu_2,\sigma_2^2) = \phi(\mu_1-\mu_2;0,\sigma_1^2+\sigma_2^2)\phi\left(u;\frac{\sigma_1^2\mu_2+\sigma_2^2\mu_1}{\sigma_1^2+\sigma_2^2},\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}\right),$$
$$\int \phi(u;\mu_1,\sigma_1^2)\phi(u;\mu_2,\sigma_2^2)du = \phi(\mu_1-\mu_2;0,\sigma_1^2+\sigma_2^2),$$
$$\int u\phi(u;\mu_1,\sigma_1^2)\phi(u;\mu_2,\sigma_2^2)du = \frac{\sigma_1^2\mu_2+\sigma_2^2\mu_1}{\sigma_1^2+\sigma_2^2}\phi(\mu_1-\mu_2;0,\sigma_1^2+\sigma_2^2),$$
$$\int u^2\phi(u;\mu_1,\sigma_1^2)\phi(u;\mu_2,\sigma_2^2)du = \left[\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}+\left(\frac{\sigma_1^2\mu_2+\sigma_2^2\mu_1}{\sigma_1^2+\sigma_2^2}\right)^2\right]\phi(\mu_1-\mu_2;0,\sigma_1^2+\sigma_2^2).$$

Then using the above two lemmas, for the components $\tilde{S}_{jn}(x)$ or $\tilde{s}_j$ for $j=0,1,2$, $\tilde{T}_{ln}(x)$ for $l=0,1$, in the definition of $\hat{g}_n(x)$ given in (3.3.8), we can get the following series of results on the asymptotic expansions of their expectations and variances. For brevity, in the proof, we denote $\delta_{jh}^2 = h^2+(\lambda+j)\sigma_u^2$ for $j=0,1,2$.

Lemma 9. *For $\tilde{S}_{n0}(x)$, we have*

$$E(\tilde{S}_{n0}(x)) = f_{0,\lambda}(x) + h^2 f_{0,\lambda}''(x)/2 + o(h^2), \quad \lambda \geq 0,$$
$$Var(\tilde{S}_{n0}(x)) = \begin{cases} \dfrac{f_{0,\lambda/2}(x)}{2n\sqrt{\pi\lambda\sigma_u^2}} - \dfrac{f_{0,\lambda}^2(x)}{n} + O\left(\dfrac{h^2}{n}\right), & \lambda > 0, \\ \dfrac{f_{0,0}(x)}{2nh\sqrt{\pi}} - \dfrac{f_{0,0}^2(x)}{n} + O\left(\dfrac{h}{n}\right), & \lambda = 0. \end{cases}$$

Proof of Lemma 9. By the independence of X and U, and applying Lemma 7 with $a=1$, $c=(\lambda+1)\sigma_u^2$ and $m(t)=f_X(t)$, for $\tilde{S}_{0n}$, we have

$$E\phi(x,Z,\delta_{0h}^2) = \iint \phi(x;t+u,\delta_{0h}^2)\phi(u;0,\sigma_u^2)f_X(t)dudt$$
$$= \int \phi(t-x;0,(\lambda+1)\sigma_u^2)f_X(t)dt + \frac{h^2}{2}\int f_X''(t)\phi(t-x;0,(\lambda+1)\sigma_u^2)dt + o(h^2)$$
$$= f_{0,\lambda}(x) + h^2 f_{0,\lambda}''(x)/2 + o(h^2).$$

Also, applying Lemma (7) with $a = 1/2$, $c = (\lambda+2)\sigma_u^2/2$, and $m(t) = f_X(t)$, we have

$$E\phi^2(x; Z, \delta_{0h}^2) = \iint \phi^2(x; t+u, \delta_{0h}^2)\phi(u; 0, \sigma_u^2) f_X(t) du dt$$
$$= \frac{1}{2\sqrt{\pi\delta_{0h}^2}} \int \phi\left(t; x, \frac{\delta_{2h}^2}{2}\right) f_X(t)dt = \frac{f_{0,\lambda/2}(x)}{2\sqrt{\pi\delta_{0h}^2}} + \frac{h^2 f''_{0,\lambda/2}(x)}{8\sqrt{\pi\delta_{0h}^2}} + o\left(\frac{h^2}{2\sqrt{\pi\delta_{0h}^2}}\right). \text{(B.1.2)}$$

Therefore, we have

$$\text{var}\left[\tilde{S}_{0n}(x)\right] = \frac{1}{n}\left\{E\phi^2(x; Z, \delta_{0h}^2) - [E\phi(x; Z, \delta_{0h}^2)]^2\right\}$$
$$= \frac{4f_{0,\lambda/2}(x) + h^2 f''_{0,\lambda/2}(x)}{8n\sqrt{\pi\delta_{0h}^2}} - \frac{f_{0,\lambda}^2(x) + f_{0,\lambda}(x) f''_{0,\lambda}(x) h^2}{n} + o\left(\frac{h^2}{2n\sqrt{\pi\delta_{0h}^2}}\right).$$

This concludes the proof of Lemma 9. □

Lemma 10. *For $\tilde{S}_{n1}(x)$, we have*

$$E(\tilde{S}_{n1}(x)) = \frac{h^2}{(\lambda+1)\sigma_u^2} f_{1,\lambda}(x) - \frac{xh^2}{(\lambda+1)\sigma_u^2} f_{0,\lambda}(x) + o\left(h^2\right), \quad \lambda \geq 0,$$

$$Var(\tilde{S}_{n1}(x)) = \begin{cases} \frac{h^4}{2n(\lambda+2)^2\sigma_u^4\sqrt{\pi\lambda\sigma_u^2}}\left[f_{2,\lambda/2}(x) - x f_{1,\lambda/2}(x) + x^2 f_{0,\lambda/2}(x)\right] + \\ \quad \frac{h^4}{2n\lambda(\lambda+2)\sigma_u^2\sqrt{\pi\lambda\sigma_u^2}} f_{0,\lambda/2}(x) - \frac{h^4}{n(\lambda+1)^2\sigma_u^4}[f_{1,\lambda}(x) - x f_{0,\lambda}(x)]^2, \quad \lambda > 0 \\ \frac{h}{4n\sqrt{\pi}} f_{0,0}(x) + o\left(\frac{h}{n}\right), \quad \lambda = 0. \end{cases}$$

Proof of Lemma 10. Applying Lemma 7 with $a = 1$, $c = (\lambda+1)\sigma_u^2$, $m(t) = f_X(t)$ and $t f_X(t)$, and Lemma 8, we have

$$E(Z-x)\phi(x, Z, \delta_{0h}^2) = \iint (t+u-x)\phi(x; t+u, \delta_{0h}^2)\phi(u; 0, \sigma_u^2) f_X(t) du dt$$
$$= \int (t-x)\left[\int \phi(u; x-t, \delta_{0h}^2)\phi(u, 0, \sigma_u^2) du\right] f_X(t) dt$$
$$+ \int \left[\int u\phi(u; x-t, \lambda\sigma_u^2)\phi(u; 0, \sigma_u^2) du\right] f_X(t) dt$$

$$= \int \phi(t-x;0,\delta_{1h}^2)(t-x)f_X(t)dt - \frac{\sigma_u^2}{\delta_{1h}^2}\int \phi(t-x;0,\delta_{1h}^2)(t-x)f_X(t)dt$$
$$= \frac{\delta_{0h}^2}{\delta_{1h}^2}\int \phi(t-x;0,\delta_{1h}^2)tf_X(t)dt - \frac{\delta_{0h}^2 x}{\delta_{1h}^2}\int \phi(t-x;0,\delta_{1h}^2)f_X(t)dt$$
$$= \frac{\delta_{0h}^2}{\delta_{1h}^2}\left[f_{1,\lambda}(x)+\frac{h^2}{2}f''_{1,\lambda}(x)\right] - \frac{x\delta_{0h}^2}{\delta_{1h}^2}\left[f_{0,\lambda}(x)+\frac{h^2}{2}f''_{0,\lambda}(x)\right] + o\left(\frac{\delta_{0h}^2 h^2}{2\delta_{1h}^2}\right).$$

We also have

$$E(Z-x)^2\phi^2(x,Z,\delta_{0h}^2) = \iint (t+u-x)^2\phi^2(x;t+u,\delta_{0h}^2)\phi(u;0,\sigma_u^2)f_X(t)dudt$$
$$= \frac{1}{2\sqrt{\pi\delta_{0h}^2}}\iint (t+u-x)^2\phi\left(u;x-t,\frac{\delta_{0h}^2}{2}\right)\phi(u;0,\sigma_u^2)f_X(t)dudt$$
$$= \frac{1}{2\sqrt{\pi\delta_{0h}^2}}\iint (t+u-x)^2\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)\phi\left(u;\frac{2\sigma_u^2(x-t)}{\delta_{2h}^2},\frac{\sigma_u^2\delta_{0h}^2}{\delta_{2h}^2}\right)f_X(t)dudt$$
$$= \left(\frac{1}{2\sqrt{\pi\delta_{0h}^2}} - \frac{1}{\sqrt{\pi\delta_{0h}^2}}\cdot\frac{2\sigma_u^2}{\delta_{2h}^2}\right)\int (t-x)^2\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt$$
$$+\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\int\left(\frac{\sigma_u^2\delta_{0h}^2}{\delta_{2h}^2}+\left[\frac{2\sigma_u^2(x-t)}{\delta_{2h}^2}\right]^2\right)\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt$$
$$= \frac{1}{2\sqrt{\pi\delta_{0h}^2}}\left(1-\frac{2\sigma_u^2}{\delta_{2h}^2}\right)^2\int (t-x)^2\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt + \frac{\sigma_u^2\delta_{0h}}{2\delta_{2h}^2\sqrt{\pi}}\int\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt$$
$$= \frac{\delta_{0h}^3}{2\delta_{2h}^4\sqrt{\pi}}\int (t-x)^2\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt + \frac{\sigma_u^2\delta_{0h}^2}{2\delta_{2h}^2\sqrt{\pi\delta_{0h}^2}}\int\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt$$
$$= \frac{\delta_{0h}^3}{2\delta_{2h}^4\sqrt{\pi}}\left(\left[f_{2,\lambda/2}(x)+\frac{h^2}{4}f''_{2,\lambda/2}(x)+o(h^2)\right] - x\left[f_{1,\lambda/2}(x)+\frac{h^2}{4}f''_{1,\lambda/2}(x)+o(h^2)\right]\right)$$
$$+\left(\frac{x^2\delta_{0h}^3}{2\delta_{2h}^4\sqrt{\pi}}+\frac{\sigma_u^2\delta_{0h}}{2\delta_{2h}^2\sqrt{\pi}}\right)\left[f_{0,\lambda/2}(x)+\frac{h^2}{4}f''_{0,\lambda/2}(x)+o(h^2)\right].$$

Therefore, we have

$$\mathrm{var}\left[\tilde{S}_{1n}(x)\right] = \frac{h^4\left\{E(Z-x)^2\phi^2(x;Z,\delta_{0h}^2) - [E(Z-x)\phi(x;Z,\delta_{0h}^2)]^2\right\}}{n\delta_{0h}^4} =$$
$$\frac{h^4}{2n\delta_{2h}^4\sqrt{\pi\delta_{0h}^2}}\left(\left[f_{2,\lambda/2}(x)+\frac{h^2}{4}f''_{2,\lambda/2}(x)+o(h^2)\right] - x\left[f_{1,\lambda/2}(x)+\frac{h^2}{4}f''_{1,\lambda/2}(x)+o(h^2)\right]\right)$$
$$+\left(\frac{h^4x^2}{2\delta_{2h}^4 n\sqrt{\pi\delta_{0h}^2}}+\frac{\sigma_u^2h^4}{2n\delta_{2h}^2\sqrt{\pi}\delta_{0h}^3}\right)\left[f_{0,\lambda/2}(x)+\frac{h^2}{4}f''_{0,\lambda/2}(x)+o(h^2)\right]$$

$$-\frac{1}{n}\left(\frac{h^2}{\delta_{1h}^2}\left[f_{1,\lambda}(x)+\frac{h^2}{2}f''_{1,\lambda}(x)\right]-\frac{xh^2}{\delta_{1h}^2}\left[f_{0,\lambda}(x)+\frac{h^2}{2}f''_{0,\lambda}(x)\right]+o\left(\frac{h^2}{2\delta_{1h}^2}\right)\right)^2.$$

This concludes the proof of Lemma 10. □

Lemma 11. *For $\tilde{S}_{n2}(x)$, we have*

$$E(\tilde{S}_{n2}(x))=h^2f_{0,\lambda}(x)+o(h^2),\quad \lambda\geq 0,$$

$$Var(\tilde{S}_{n2}(x))=\begin{cases}\frac{h^4}{n}\left[\frac{f_{0,\lambda/2}(x)}{2\sqrt{\pi\lambda\sigma_u^2}}-f_{0,\lambda}^2(x)\right]+o\left(\frac{h^4}{n}\right), & \lambda>0,\\ \frac{3h^3f_{0,0}(x)}{8n\sqrt{\pi}}+o\left(\frac{h^3}{n}\right), & \lambda=0.\end{cases}$$

Proof of Lemma 11. Note that

$$E(Z-x)^2\phi(x,Z,\delta_{0h}^2)=\iint(t+u-x)^2\phi(x;t+u,\delta_{0h}^2)\phi(u;0,\sigma_u^2)f_X(t)dudt$$
$$=\iint(t+u-x)^2\phi\left(t,x,\delta_{1h}^2\right)\phi\left(u;\frac{\sigma_u^2(x-t)}{\delta_{1h}^2},\frac{\sigma_u^2\delta_{0h}^2}{\delta_{1h}^2}\right)f_X(t)dudt$$
$$=\int(t-x)^2\phi\left(t,x,\delta_{1h}^2\right)f_X(t)dt-\frac{2\sigma_u^2}{\delta_{1h}^2}\int(t-x)^2\phi\left(t,x,\delta_{1h}^2\right)f_X(t)dt$$
$$+\int\left(\frac{\sigma_u^2\delta_{0h}^2}{\delta_{1h}^2}+\left[\frac{\sigma_u^2(x-t)}{\delta_{1h}^2}\right]^2\right)\phi\left(t,x,\delta_{1h}^2\right)f_X(t)dt$$
$$=\frac{\delta_{0h}^4}{\delta_{1h}^4}\int(t-x)^2\phi\left(t,x,\delta_{1h}^2\right)f_X(t)dt+\frac{\sigma_u^2\delta_{0h}^2}{\delta_{1h}^2}\int\phi\left(t,x,\delta_{1h}^2\right)f_X(t)dt$$
$$=\frac{\delta_{0h}^4}{\delta_{1h}^4}\left[f_{2,\lambda}(x)+\frac{h^2}{2}f''_{2,\lambda}(x)+o(h^2)-2x\left(f_{1,\lambda}(x)+\frac{h^2}{2}f''_{1,\lambda}(x)+o(h^2)\right)\right.$$
$$\left.+x^2\left(f_{0,\lambda}(x)+\frac{h^2}{2}f''_{0,\lambda}(x)+o(h^2)\right)\right]+\frac{\sigma_u^2\delta_{0h}^2}{\delta_{1h}^2}\left[f_{0,\lambda}(x)+\frac{h^2}{2}f''_{0,\lambda}(x)+o(h^2)\right].$$

Therefore, we get

$$E[\tilde{S}_{2n}(x)]=\frac{h^4}{\delta_{0h}^4}\left(\left[\frac{\delta_{0h}^2}{\delta_{1h}^2}\right]^2\left[f_{2,\lambda}(x)+\frac{h^2}{2}f''_{2,\lambda}(x)+o(h^2)-2x\left(f_{1,\lambda}(x)+\frac{h^2}{2}f''_{1,\lambda}(x)+o(h^2)\right)\right.\right.$$
$$\left.\left.+x^2\left(f_{0,\lambda}(x)+\frac{h^2}{2}f''_{0,\lambda}(x)+o(h^2)\right)\right]+\frac{\sigma_u^2\delta_{0h}^2}{\delta_{1h}^2}\left[f_{0,\lambda}(x)+\frac{h^2}{2}f''_{0,\lambda}(x)+o(h^2)\right]\right)$$

$$+\frac{\lambda\sigma_u^2 h^2}{\delta_{0h}^2}\left[f_{0,\lambda}(x)+\frac{h^2}{2}f''_{0,\lambda}(x)+o(h^2)\right].$$

Then we have to calculate $E(Z-x)^4\phi^2(x,Z,\delta_{0h}^2)$. Recall that for $u\sim N(\mu,\sigma_u^2)$, we have $Eu^3=3\mu\sigma_u^2+\mu^3$, $Eu^4=3\sigma^4+6\mu^2\sigma_u^2+\mu^4$. So,

$$E(Z-x)^4\phi^2(x,Z,\delta_{0h}^2)=\iint(t+u-x)^4\phi^2(u;x-t,\delta_{0h}^2)\phi(u;0,\sigma_u^2)f_X(t)dudt$$

$$=\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\iint(t+u-x)^4\phi\left(u;x-t,\frac{\delta_{0h}^2}{2}\right)\phi(u;0,\sigma_u^2)f_X(t)dudt$$

$$=\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\iint(t+u-x)^4\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)\phi\left(u;\frac{2\sigma_u^2(x-t)}{\delta_{2h}^2},\frac{\sigma_u^2\delta_{0h}^2}{\delta_{2h}^2}\right)f_X(t)dudt$$

$$=\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\int(t-x)^4\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt-\frac{4\sigma_u^2}{\delta_{2h}^2\sqrt{\pi\delta_{0h}^2}}\int(t-x)^4\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt$$

$$+\frac{6}{2\sqrt{\pi\delta_{0h}^2}}\int(t-x)^2\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)\left[\left(\frac{2\sigma_u^2(x-t)}{\delta_{2h}^2}\right)^2+\frac{\sigma_u^2\delta_{0h}^2}{\delta_{2h}^2}\right]f_X(t)dt$$

$$+\frac{4}{2\sqrt{\pi\delta_{0h}^2}}\int(t-x)\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)\left[3\frac{\sigma_u^2\delta_{0h}^2}{\delta_{2h}^2}\frac{2\sigma_u^2(x-t)}{\delta_{2h}^2}+\left[\frac{2\sigma_u^2(x-t)}{\delta_{2h}^2}\right]^3\right]f_X(t)dt$$

$$+\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\int\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)\left[3\left(\frac{\sigma_u^2\delta_{0h}^2}{\delta_{2h}^2}+\left(\frac{2\sigma_u^2(x-t)}{\delta_{2h}^2}\right)^2\right)^2-2\left(\frac{2\sigma_u^2(x-t)}{\delta_{2h}^2}\right)^4\right]dt$$

It can be further written as

$$\left(\frac{1}{2\sqrt{\pi\delta_{0h}^2}}-\frac{4\sigma_u^2}{\delta_{2h}^2\sqrt{\pi\delta_{0h}^2}}\right)\int(t-x)^4\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt$$

$$+\frac{6}{2\sqrt{\pi\delta_{0h}^2}}\int(t-x)^2\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)\left[\left(\frac{2\sigma_u^2(x-t)}{\delta_{2h}^2}\right)^2+\frac{\sigma_u^2\delta_{0h}^2}{\delta_{2h}^2}\right]f_X(t)dt$$

$$+\frac{4}{2\sqrt{\pi\delta_{0h}^2}}\int(t-x)\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)\left[3\frac{\sigma_u^2\delta_{0h}^2}{\delta_{2h}^2}\frac{2\sigma_u^2(x-t)}{\delta_{2h}^2}+\left[\frac{2\sigma_u^2(x-t)}{\delta_{2h}^2}\right]^3\right]f_X(t)dt$$

$$+\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\int\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)\left[3\left(\frac{\sigma_u^2\delta_{0h}^2}{\delta_{2h}^2}+\left(\frac{2\sigma_u^2(x-t)}{\delta_{2h}^2}\right)^2\right)^2-2\left(\frac{2\sigma_u^2(x-t)}{\delta_{2h}^2}\right)^4\right]dt$$

Let

$$A(h;\lambda) = \frac{1}{2\sqrt{\pi\delta_{0h}^2}}\left[\frac{\delta_{0h}^2}{\delta_{2h}^2}\right]^4, \quad B(h;\lambda) = \frac{3\sigma_u^2\sqrt{\delta_{0h}^2}}{\delta_{2h}^2}\left[1 - \frac{4\sigma_u^2}{\delta_{2h}^2} + \frac{4\sigma_u^4}{\delta_{2h}^4}\right] = \frac{3\sigma_u^2\delta_{0h}^5}{\sqrt{\pi}\delta_{2h}^6},$$
$$C(h;\lambda) = \frac{3\sigma_u^4\delta_{0h}^2\sqrt{\delta_{0h}^2}}{2\sqrt{\pi}\delta_{2h}^4}.$$

Then,

$$\begin{aligned}
E(Z-x)^4\phi^2(x,Z,\delta_{0h}^2) &= A(h;\lambda)\int (t-x)^4\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt \\
&\quad +B(h;\lambda)\int (t-x)^2\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt + C(h;\lambda)\int \phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt \\
&= A(h;\lambda)\int t^4\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt - 4xA(h;\lambda)\int t^3\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt \\
&\quad +(6x^2A(h;\lambda)+B(h;\lambda))\int t^2\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt \\
&\quad -(4x^3A(h;\lambda)+2xB(h;\lambda))\int t\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt \\
&\quad +[x^4A(h;\lambda)+x^2B(h;\lambda)+C(h;\lambda)]\int \phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt \\
&= A(h;\lambda)\left[f_{4,\lambda/2}(x) + \frac{h^2}{4}f''_{4,\lambda/2}(x) + o(h^2)\right] - 4xA(h;\lambda)\left[f_{3,\lambda/2}(x) + \frac{h^2}{4}f''_{3,\lambda/2}(x) + o(h^2)\right] \\
&\quad +(6x^2A(h;\lambda)+B(h;\lambda))\left[f_{2,\lambda/2}(x) + \frac{h^2}{4}f''_{2,\lambda/2}(x) + o(h^2)\right] \\
&\quad -(4x^3A(h;\lambda)+2xB(h;\lambda))\left[f_{1,\lambda/2}(x) + \frac{h^2}{4}f''_{1,\lambda/2}(x) + o(h^2)\right] \\
&\quad +[x^4A(h;\lambda)+x^2B(h;\lambda)+C(h;\lambda)]\left[f_{0,\lambda/2}(x) + \frac{h^2}{4}f''_{0,\lambda/2}(x) + o(h^2)\right].
\end{aligned}$$

Summarizing above derivations eventually leads to

$$\begin{aligned}
\text{var}[\tilde{S}_{2n}(x)] &= \frac{h^8}{n\delta_{0h}^8}\left[E(Z-x)^4\phi^2(x;Z,\delta_{0h}^2) - (E(Z-x)^2\phi(x;Z,\delta_{0h}^2))^2\right] \\
&\quad +\frac{\lambda^2\sigma_u^4h^4}{n\delta_{0h}^4}\left[E\phi^2(x;Z,\delta_{0h}^2) - (E\phi(x;Z,\delta_{0h}^2))^2\right] + \frac{2\lambda\sigma_u^2h^6}{n\delta_{0h}^6}\cdot \\
&\qquad \left[E(Z-x)^2\phi^2(x;Z,\delta_{0h}^2) - E(Z-x)^2\phi(x;Z,\delta_{0h}^2)E\phi(x;Z,\delta_{0h}^2)\right].
\end{aligned}$$

This concludes the proof of Lemma 11. □

Lemma 12. *For $\tilde{T}_{n0}(x)$, we have*

$$
\begin{aligned}
E(\tilde{T}_{n0}(x)) &= g_{0,\lambda}(x) + \frac{h^2}{2} g''_{0,\lambda}(x) + o(h^2), \quad \lambda > 0, \\
Var(\tilde{T}_{n0}(x)) &= \begin{cases} \frac{1}{n}\left[\frac{1}{2\sqrt{\lambda\pi\sigma_u^2}}[G_{0,\lambda/2}(x) + H_{0,\lambda/2}(x)] - g_{0,\lambda}^2(x)\right] + O\left(\frac{h^2}{n}\right), & \lambda > 0, \\ \frac{1}{2nh\sqrt{\pi}}[G_{0,0}(x) + H_{0,0}(x)] + o\left(\frac{1}{nh}\right), & \lambda = 0. \end{cases}
\end{aligned}
$$

Proof of Lemma 12. Note that

$$
\begin{aligned}
& E[Y\phi(x; Z, \delta_{0h}^2)] = E[(g(X) + \varepsilon)\phi(x; Z, \delta_{0h}^2)] \\
&= E\left(E[(g(X) + \varepsilon)\phi(x; Z, \delta_{0h}^2)\Big| X, U]\right) = E[g(X)\phi(x; Z, \delta_{0h}^2)] \\
&= \int g(t) f_X(t)\phi(x - t; 0, (\lambda + 1)\sigma^2)dt + \frac{h^2}{2}\frac{\partial^2}{\partial x^2}\int g(t) f_X(t)\phi(x - t; 0, (\lambda + 1)\sigma_u^2)dt + o(h^2).
\end{aligned}
$$

Note that $\tau^2(X) = E(\varepsilon^2|X)$, we also have

$$
\begin{aligned}
& E[Y^2\phi^2(x; Z, \delta_{0h}^2)] = E\left(E[(g(X) + \varepsilon)^2\phi^2(x; Z, \delta_{0h}^2)\Big| X, U]\right) \\
&= E[(g^2(X) + \tau^2(X))\phi^2(x; Z, \delta_{0h}^2)] = \int [g^2(t) + \tau^2(t)]\phi^2(u; x - t, \delta_{0h}^2)\phi(u; 0, \sigma_u^2) f_X(t)dt \\
&= \frac{1}{2\sqrt{\pi\delta_{0h}^2}}\int [g^2(t) + \tau^2(t)]\phi(x - t; 0, (\lambda + 2)\sigma_u^2/2) f_X(t)dt \\
&\quad + \frac{h^2}{8\sqrt{\pi\delta_{0h}^2}}\frac{\partial^2}{\partial x^2}\int [g^2(t) + \tau^2(t)]\phi(x - t; 0, (\lambda + 2)\sigma_u^2/2) f_X(t)dt + o\left(\frac{h^2}{2\sqrt{\pi\delta_{0h}^2}}\right).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\text{var}\left[\tilde{T}_{0n}(x)\right] &= \frac{1}{n}\left\{E[Y^2\phi^2(x; Z, \delta_{0h}^2)] - (E[Y\phi(x; Z, \delta_{0h}^2)])^2\right\} \\
&= \frac{1}{2n\sqrt{\pi\delta_{0h}^2}}\int [g^2(t) + \tau^2(t)]\phi(x - t; 0, (\lambda + 2)\sigma_u^2/2) f_X(t)dt \\
&\quad + \frac{h^2}{8n\sqrt{\pi\delta_{0h}^2}}\frac{\partial^2}{\partial x^2}\int [g^2(t) + \tau^2(t)]\phi(x - t; 0, (\lambda + 2)\sigma_u^2/2) f_X(t)dt
\end{aligned}
$$

$$+o\left(\frac{h^2}{2n\sqrt{\pi\delta_{0h}^2}}\right)-\frac{1}{n}\Bigg[\int g(t)f_X(t)\phi(x-t;0,(\lambda+1)\sigma_u^2)dt$$

$$+\frac{h^2}{2}\frac{\partial^2}{\partial x^2}\int g(t)f_X(t)\phi(x-t;0,(\lambda+1)\sigma_u^2)dt+o(h^2)\Bigg]^2.$$

This implies the result in Lemma 12. □

Lemma 13. *For* $\tilde{T}_{n1}(x)$*, we have*

$$E(\tilde{T}_{n1}(x))=\frac{h^2}{(\lambda+1)\sigma_u^2}g_{1,\lambda}(x)-\frac{xh^2}{(\lambda+1)\sigma_u^2}g_{0,\lambda}(x)+o\left(h^2\right),\quad \lambda\geq 0,$$

$$Var(\tilde{T}_{n1}(x))=$$

$$\begin{cases}\frac{h^4}{n(\lambda+2)^2\sqrt{\lambda\pi}\sigma_u^5}\left[\frac{1}{2}[G_{2,\lambda/2}(x)+H_{2,\lambda/2}(x)]-x[G_{1,\lambda/2}(x)+H_{1,\lambda/2}(x)]\right]+\\ \qquad\frac{h^4}{n(\lambda+2)\sqrt{\lambda\pi}\sigma_u^3}\left[\frac{x^2}{(\lambda+2)\sigma_u^2}+\frac{1}{\lambda}\right][G_{0,\lambda/2}(x)+H_{0,\lambda/2}(x)]-\\ \qquad\frac{h^4}{n(\lambda+1)^2\sigma_u^4}[g_{1,\lambda}(x)-xg_{0,\lambda}(x)]^2,\quad \lambda>0,\\ \frac{h}{4n\sqrt{\pi}}\left[G_{0,0}(x)+H_{0,0}(x)\right]+o\left(\frac{h}{n}\right),\quad \lambda=0.\end{cases}$$

Proof of Lemma 13. Note that

$$\begin{aligned}
&E\left[Y(Z-x)\phi(x;Z,\delta_{0h}^2)\right]=E\left[(g(X)+\varepsilon)(Z-x)\phi(x;Z,\delta_{0h}^2)\right]\\
&=E\left(E[(g(X)+\varepsilon)(Z-x)\phi(x;Z,\delta_{0h}^2)\Big|X,U]\right)=E\left[g(X)(Z-x)\phi(x;Z,\delta_{0h}^2)\right]\\
&=\iint g(t)(t+u-x)\phi(x;t+u,\delta_{0h}^2)\phi(u;0,\sigma_u^2)f_X(t)dudt\\
&=\int g(t)(t-x)\left[\int\phi(u;x-t,\delta_{0h}^2)\phi(u,0,\sigma_u^2)du\right]f_X(t)dt\\
&\quad+\int g(t)\left[\int u\phi(u;x-t,\lambda\sigma_u^2)\phi(u;0,\sigma_u^2)du\right]f_X(t)dt\\
&=\int\phi(t,x,\delta_{1h}^2)g(t)(t-x)f_X(t)dt-\frac{\sigma_u^2}{\delta_{1h}^2}\int\phi(t,x,\delta_{1h}^2)g(t)(t-x)f_X(t)dt\\
&=\frac{\delta_{0h}^2}{\delta_{1h}^2}\int\phi(t,x,\delta_{1h}^2)tg(t)f_X(t)dt-\frac{\delta_{0h}^2x}{\delta_{1h}^2}\int\phi(t,x,\delta_{1h}^2)g(t)f_X(t)dt
\end{aligned}$$

$$= \frac{\delta_{0h}^2}{\delta_{1h}^2}\left[\int tg(t)f_X(t)\phi(t;x,(\lambda+1)\sigma_u^2)dt + \frac{h^2}{2}\frac{\partial^2}{\partial x^2}\int tg(t)f_X(t)\phi(t;x,(\lambda+1)\sigma_u^2)dt + o(h^2)\right]$$
$$-\frac{\delta_{0h}^2 x}{\delta_{1h}^2}\left[\int g(t)f_X(t)\phi(t;x,(\lambda+1)\sigma_u^2)dt + \frac{h^2}{2}\frac{\partial^2}{\partial x^2}\int g(t)f_X(t)\phi(t;x,(\lambda+1)\sigma_u^2)dt + o(h^2)\right]$$

Next, we see that

$$E[Y^2(Z-x)^2\phi^2(x;Z,\delta_{0h}^2)] = E\left(E[(g(X)+\varepsilon)^2(Z-x)^2\phi^2(x;Z,\delta_{0h}^2)\Big|X,U]\right)$$
$$= E[(g^2(X)+\tau^2(X))(Z-x)^2\phi^2(x;Z,\delta_{0h}^2)]$$
$$= \int [g^2(t)+\tau^2(t)]\int (t+u-x)^2\phi^2(x;t+u,\delta_{0h}^2)\phi(u;0,\sigma_u^2)f_X(t)dudt$$
$$= \frac{\delta_{0h}^3}{2\delta_{2h}^4\sqrt{\pi}}\left[\int [g^2(t)+\tau^2(t)]t^2 f_X(t)\phi(x-t;0,(\lambda+2)\sigma_u^2/2)dt + o(h^2)\right]$$
$$+\frac{h^2\delta_{0h}^3}{8\delta_{2h}^4\sqrt{\pi}}\frac{\partial^2}{\partial x^2}\int [g^2(t)+\tau^2(t)]t^2 f_X(t)\phi(x-t;0,(\lambda+2)\sigma_u^2/2)dt$$
$$-\frac{x\delta_{0h}^3}{\delta_{2h}^4\sqrt{\pi}}\left[\int [g^2(t)+\tau^2(t)]t f_X(t)\phi(x-t;0,(\lambda+2)\sigma_u^2/2)dt + o(h^2)\right]$$
$$-\frac{xh^2\delta_{0h}^3}{4\delta_{2h}^4\sqrt{\pi}}\frac{\partial^2}{\partial x^2}\int [g^2(t)+\tau^2(t)]t f_X(t)\phi(x-t;0,(\lambda+2)\sigma_u^2/2)dt$$
$$+\frac{x^2\delta_{0h}^3}{2\delta_{2h}^4\sqrt{\pi}}\left[\int [g^2(t)+\tau^2(t)] f_X(t)\phi(x-t;0,(\lambda+2)\sigma_u^2/2)dt + o(h^2)\right]$$
$$+\frac{x^2h^2\delta_{0h}^3}{8\delta_{2h}^4\sqrt{\pi}}\frac{\partial^2}{\partial x^2}\int [g^2(t)+\tau^2(t)] f_X(t)\phi(x-t;0,(\lambda+2)\sigma_u^2/2)dt$$
$$+\frac{\sigma_u^2\delta_{0h}}{2\delta_{2h}^2\sqrt{\pi}}\int [g^2(t)+\tau^2(t)] f_X(t)\phi(x-t;0,(\lambda+2)\sigma_u^2/2)dt$$
$$+\frac{\sigma_u^2h^2\delta_{0h}}{8\delta_{2h}^2\sqrt{\pi}}\frac{\partial^2}{\partial x^2}\int [g^2(t)+\tau^2(t)] f_X(t)\phi(x-t;0,(\lambda+2)\sigma_u^2/2)dt + o\left(\frac{h^2\delta_{0h}}{2\delta_{2h}^2\sqrt{\pi}}\right)$$

Therefore,

$$\mathrm{var}\left[\tilde{T}_{1n}(x)\right] = \frac{h^4}{n\delta_{0h}^4}\left\{E[Y^2(Z-x)^2\phi^2(x;Z,\delta_{0h}^2)] - \left(E[Y(Z-x)\phi(x;Z,\delta_{0h}^2)]\right)^2\right\}$$
$$= \frac{h^4}{2n\delta_{2h}^4\sqrt{\pi\delta_{0h}^2}}\left[\int [g^2(t)+\tau^2(t)]t^2 f_X(t)\phi(x-t;0,(\lambda+2)\sigma_u^2/2)dt + o(h^2)\right]$$
$$+\frac{h^6}{8n\delta_{2h}^4\sqrt{\pi\delta_{0h}^2}}\frac{\partial^2}{\partial x^2}\int [g^2(t)+\tau^2(t)]t^2 f_X(t)\phi(x-t;0,(\lambda+2)\sigma_u^2/2)dt$$

$$
\begin{aligned}
&-\frac{xh^4}{n\delta_{2h}^4\sqrt{\pi\delta_{0h}^2}}\left[\int[g^2(t)+\tau^2(t)]tf_X(t)\phi(x-t;0,(\lambda+2)\sigma_u^2/2)dt+o(h^2)\right] \\
&-\frac{xh^6}{4n\delta_{2h}^4\sqrt{\pi\delta_{0h}^2}}\frac{\partial^2}{\partial x^2}\int[g^2(t)+\tau^2(t)]tf_X(t)\phi(x-t;0,(\lambda+2)\sigma_u^2/2)dt \\
&+\frac{x^2h^4}{2n\delta_{2h}^4\sqrt{\pi\delta_{0h}^2}}\left[\int[g^2(t)+\tau^2(t)]f_X(t)\phi(x-t;0,(\lambda+2)\sigma_u^2/2)dt+o(h^2)\right] \\
&+\frac{x^2h^6}{8n\delta_{2h}^4\sqrt{\pi\delta_{0h}^2}}\frac{\partial^2}{\partial x^2}\int[g^2(t)+\tau^2(t)]f_X(t)\phi(x-t;0,(\lambda+2)\sigma_u^2/2)dt \\
&+\frac{\sigma_u^2h^4}{2n\delta_{2h}^2\delta_{0h}^3\sqrt{\pi}}\left[\int[g^2(t)+\tau^2(t)]f_X(t)\phi(x-t;0,(\lambda+2)\sigma_u^2/2)dt+o(h^2)\right] \\
&+\frac{\sigma_u^2h^6}{8n\delta_{2h}^2\delta_{0h}^3\sqrt{\pi}}\frac{\partial^2}{\partial x^2}\int[g^2(t)+\tau^2(t)]f_X(t)\phi(x-t;0,(\lambda+2)\sigma_u^2/2)dt \\
&-\frac{h^4}{n\delta_{1h}^4}\Bigg[\int tg(t)f_X(t)\phi(x-t;0,(\lambda+1)\sigma_u^2)dt+\frac{h^2}{2}\frac{\partial^2}{\partial x^2}\int tg(t)f_X(t)\phi(x-t;0,(\lambda+1)\sigma_u^2)dt \\
&-x\int g(t)f_X(t)\phi(t;x,(\lambda+1)\sigma_u^2)dt-\frac{xh^2}{2}\frac{\partial^2}{\partial x^2}\int g(t)f_X(t)\phi(t;x,(\lambda+1)\sigma_u^2)dt+o(h^2)\Bigg]^2.
\end{aligned}
$$

This concludes the proof of Lemma 13. □

Proof of Theorem 5. To verify the Lyapunov condition, we have to find out the asymptotic expansions of $Ev^2(x)$, and an upper bound for $E|v^3(x)|$. Note that

$$
\begin{aligned}
Ev^2(x) = {} & c_{0\lambda}^2E\xi_{0\lambda}^2(x)+c_{1\lambda}^2E\xi_{1\lambda}^2(x)+c_{2\lambda}^2E\xi_{2\lambda}^2(x)+d_{0\lambda}^2E\eta_{0\lambda}^2(x)+d_{1\lambda}^2E\eta_{1\lambda}^2(x) \\
&+2c_{0\lambda}c_{1\lambda}E[\xi_{0\lambda}(x)\xi_{1\lambda}(x)]+2c_{0\lambda}c_{2\lambda}E[\xi_{0\lambda}(x)\xi_{2\lambda}(x)]+2c_{0\lambda}d_{0\lambda}E[\xi_{0\lambda}(x)\eta_{0\lambda}(x)] \\
&+2c_{0\lambda}d_{1\lambda}E[\xi_{0\lambda}(x)\eta_{1\lambda}(x)]+2c_{1\lambda}c_{2\lambda}E[\xi_{1\lambda}(x)\xi_{2\lambda}(x)]+2c_{1\lambda}d_{0\lambda}E[\xi_{1\lambda}(x)\eta_{0\lambda}(x)] \\
&+2c_{1\lambda}d_{1\lambda}E[\xi_{1\lambda}(x)\eta_{1\lambda}(x)]+2c_{2\lambda}d_{0\lambda}E[\xi_{2\lambda}(x)\eta_{0\lambda}(x)]+2c_{2\lambda}d_{1\lambda}E[\xi_{2\lambda}(x)\eta_{1\lambda}(x)] \\
&+2d_{0\lambda}d_{1\lambda}E[\eta_{0\lambda}(x)\eta_{1\lambda}(x)].
\end{aligned}
$$

Routine and tedious calculations show that when $\lambda > 0$, except for

$$
E[\xi_{0\lambda}(x)\eta_{0\lambda}(x)] = \frac{1}{2\sqrt{\pi\lambda\sigma_u^2}}g_{0,\lambda/2}(x) - g_{0,\lambda}(x)f_{0,\lambda}(x) + O(h^2),
$$

all other expectations of the cross products are of the order $O(h^2)$, which, together with the previous derivations with respect to $E\xi_{0\lambda}^2(x)$, $E\xi_{1\lambda}^2(x)$, $E\xi_{2\lambda}^2(x)$, $E\eta_{0\lambda}^2(x)$, $E\eta_{1\lambda}^2(x)$, we can obtain

$$\begin{aligned} Ev^2(x) = {} & c_{0\lambda}^2 \left[\frac{1}{2\sqrt{\pi\lambda\sigma_u^2}} f_{0,\lambda/2}(x) - f_{0,\lambda}^2(x) \right] + d_{0\lambda}^2 \left[\frac{1}{2\sqrt{\pi\lambda\sigma_u^2}} \{G_{0,\lambda/2}(x) + H_{0,\lambda/2}(x)\} - g_{0,\lambda}^2 \right] \\ & + 2c_{0\lambda}d_{0\lambda} \left[\frac{1}{2\sqrt{\pi\lambda\sigma_u^2}} g_{0,\lambda/2}(x) - g_{0,\lambda}(x) f_{0,\lambda}(x) \right] + O(h^2). \end{aligned}$$

When $\lambda = 0$, except for $E[\xi_{00}(x)\eta_{00}(x)] = \frac{1}{2h\sqrt{\pi}} g_{0,0}(x) + O(h)$, all other expectations of the cross products are of the order $O(h)$, which, together with the previous derivations with respect to $E\xi_{00}^2(x)$, $E\xi_{10}^2(x)$, $E\xi_{20}^2(x)$, $E\eta_{00}^2(x)$, $E\eta_{10}^2(x)$, leads to

$$\begin{aligned} Ev^2(x) &= \frac{1}{2h\sqrt{\pi}} \left[c_{00}^2 f_{00}(x) + d_{00}^2 \{G_{00}(x) + H_{00}(x)\} + 2c_{00}d_{00}g_{00}(x) \right] + O(h) \\ &= \frac{1}{2h\sqrt{\pi}} \left[\frac{G_{00}(x) + H_{00}(x)}{f_{00}^2(x)} - \frac{g_{00}^2(x)}{f_{00}^3(x)} \right] + O(h). \end{aligned}$$

To find a proper order for $E|v(x)|^3$, we have to find the orders for the expectations

$$\begin{aligned} & E(Z-x)\phi^2(x, Z, \delta_{0h}^2), \quad EY\phi^2(x, Z, \delta_{0h}^2), \quad EY(Z-x)\phi^2(x, Z, \delta_{0h}^2), \\ & EY(Z-x)^2\phi^2(x, Z, \delta_{0h}^2), \quad EY^2(Z-x)\phi^2(x, Z, \delta_{0h}^2), \\ & E\phi^3(x, Z, \delta_{0h}^2), \quad E|Z-x|^3\phi^3(x, Z, \delta_{0h}^2), \quad E|Z-x|^6\phi^3(x, Z, \delta_{0h}^2), \\ & E|Y|^3\phi^3(x, Z, \delta_{0h}^2), \quad E|Y|^3|Z-x|^3\phi^3(x, Z, \delta_{0h}^2). \end{aligned}$$

More complicated calculations show that

$$E\phi^3(x, Z, \delta_{0h}^2) = \frac{1}{2\pi\delta_{0h}^2\sqrt{3}} \left[f_{0,\lambda/3}(x) + \frac{h^2}{6} f_{0,\lambda/3}''(x) + o(h^2) \right],$$

which is $O(1)$ when $\lambda > 0$ and $O(1/h^2)$ when $\lambda = 0$, and

$$E|Z-x|^3\phi^3(x, Z, \delta_{0h}^2)$$

$$
\begin{aligned}
&\leq \frac{2}{\pi\sqrt{3}\delta_{0h}^2} \cdot \frac{8\sqrt{2}}{\sqrt{\pi}} \left(\frac{\sigma_u^2\delta_{0h}^2}{h^2+(\lambda+3)\sigma_u^2}\right)^{\frac{3}{2}} \int \phi\left(t, x, \frac{h^2+(\lambda+3)\sigma_u^2}{3}\right) f_X(t)dt \\
&\quad + \frac{2}{\pi\sqrt{3}\delta_{0h}^2} \cdot 4 \left(\frac{3\sigma_u^2}{h^2+(\lambda+3)\sigma_u^2}\right)^3 \int |x-t|^3 \phi\left(t, x, \frac{h^2+(\lambda+3)\sigma_u^2}{3}\right) f_X(t)dt \\
&\quad + \frac{2}{\pi\sqrt{3}\delta_{0h}^2} \int |x-t|^3 \phi\left(t, x, \frac{h^2+(\lambda+3)\sigma_u^2}{3}\right) f_X(t)dt
\end{aligned}
$$

which is $O(1)$ when $\lambda > 0$ and $O(1/h^2)$ when $\lambda = 0$. We also have

$$
\begin{aligned}
E|Z-x|^6\phi^3(x; Z, \delta_{0h}^2) &\leq \frac{7680\sigma_u^6}{\pi\sqrt{3}\delta_{0h}^2} \left(\frac{\sigma_u^2\delta_{0h}^2}{h^2+(\lambda+3)\sigma_u^2}\right)^3 \int \phi\left(t, x, \frac{h^2+(\lambda+3)\sigma_u^2}{3}\right) f_X(t)dt \\
&+ \frac{512}{\pi\sqrt{3}\delta_{0h}^2} \left(\frac{3\sigma_u^2}{h^2+(\lambda+3)\sigma_u^2}\right)^6 \int |x-t|^6 \phi\left(t, x, \frac{h^2+(\lambda+3)\sigma_u^2}{3}\right) f_X(t)dt \\
&+ \frac{16}{\pi\sqrt{3}\delta_{0h}^2} \int |x-t|^6 \phi\left(t, x, \frac{h^2+(\lambda+3)\sigma_u^2}{3}\right) f_X(t)dt
\end{aligned}
$$

which is $O(1)$ when $\lambda > 0$ and $O(1/h^2)$ when $\lambda = 0$. Denote $\delta(X) = E\left(|\epsilon|^3 \middle| X\right)$, then,

$$
E|Y|^3\phi^3(x; Z, \delta_{0h}^2) \leq 4E|g(X)|^3\phi^3(x; Z, \delta_{0h}^2) + 4E\delta(X)\phi^3(x; Z, \delta_{0h}^2).
$$

Eventually, we can show that

$$
E|Y|^3\phi^3(x; Z, \delta_{0h}^2) \leq \frac{2}{\pi\sqrt{3}\delta_{0h}^2} \int \left[|g(t)|^3 + \delta(t)\right] \phi\left(t; x, \frac{h^2+(\lambda+3)\sigma_u^2}{3}\right) f_X(t)dt
$$

which is $O(1)$ for $\lambda > 0$ and $O(1/h^2)$ for $\lambda = 0$. Finally, for For $E|Y|^3|Z-x|^3\phi^3(x; Z, \delta_{0h}^2)$, we can show that

$$
\begin{aligned}
&E|Y|^3|Z-x|^3\phi^3(x; Z, \delta_{0h}^2) \\
&\leq \frac{64\sqrt{2}}{\pi\sqrt{3\pi}\delta_{0h}^2} \left(\frac{\sigma_u^2\delta_{0h}^2}{h^2+(\lambda+3)\sigma_u^2}\right)^{\frac{3}{2}} \int \left[|g(t)|^3 + \delta(t)\right] \phi\left(t, x, \frac{h^2+(\lambda+3)\sigma_u^2}{3}\right) f_X(t)dt \\
&\quad + \left[\frac{32}{\pi\sqrt{3}\delta_{0h}^2} \left(\frac{3\sigma_u^2}{h^2+(\lambda+3)\sigma_u^2}\right)^3 + \frac{8}{\pi\sqrt{3}\delta_{0h}^2}\right] \cdot \\
&\quad \int \left[|g(t)|^3 + \delta(t)\right] |x-t|^3 \phi\left(t, x, \frac{h^2+(\lambda+3)\sigma_u^2}{3}\right) f_X(t)dt,
\end{aligned}
$$

which is $O(1)$ when $\lambda > 0$ and $O(1/h^2)$ when $\lambda = 0$.

Therefore, when $\lambda > 0$,

$$\frac{\sum_{i=1}^{n} E|v_{i\lambda}(x)|^3}{\left(\sum_{i=1}^{n} Ev_{i\lambda}^2(x)\right)^{3/2}} = \frac{O(n)}{O(n^{3/2})} \to 0$$

as $n \to \infty$, and when $\lambda = 0$,

$$\frac{\sum_{i=1}^{n} E|v_{i\lambda}(x)|^3}{\left(\sum_{i=1}^{n} Ev_{i\lambda}^2(x)\right)^{3/2}} = \frac{O(n/h^2)}{O((n/h)^{3/2})} = O\left(\frac{1}{\sqrt{nh}}\right) \to 0.$$

So, by Lyapunov central limit theorem, we proved Theorem 5. □

Proof of (3.3.3)-(3.3.5). By the normality assumption of V and its independence from other random variables in the model, and the kernel function K being the standard normal density, from Lemma 8, we have

$$E[K_h(Z(\lambda) - x)|Y, Z] = \int \phi(v; x - Z, h^2)\phi(v; 0, \lambda\sigma_u^2)dv = \phi(x; Z, \delta_{0h}^2),$$

which is (3.3.3). (3.3.4) can be derived from the following algebra,

$$E[(Z(\lambda) - x)K_h(Z(\lambda) - x)|Y, Z] = \int (Z + v - x)\phi(v; x - Z, h^2)\phi(v; 0, \lambda\sigma_u^2)dv$$
$$= (Z - x)\phi(x - Z; 0, \delta_{0h}^2) + \frac{\lambda\sigma_u^2(x - Z)}{\delta_{0h}^2}\phi(x - Z; 0, \delta_{0h}^2).$$

Finally, note that

$$E[(Z(\lambda) - x)^2K_h(Z(\lambda) - x)|Y, Z] = \int (Z + v - x)^2\phi(v; x - Z, h^2)\phi(v; 0, \lambda\sigma_u^2)dv$$
$$= (Z - x)^2 \int \phi(v; x - Z, h^2)\phi(v; 0, \lambda\sigma_u^2)dv + 2(Z - x)\int v\phi(v; x - Z, h^2)\phi(v; 0, \lambda\sigma_u^2)dv$$
$$+ \int v^2\phi(v; x - Z, h^2)\phi(v; 0, \lambda\sigma_u^2)dv$$
$$= (Z - x)^2\phi(x - Z; 0, \delta_{0h}^2) - \frac{2\lambda\sigma_u^2(Z - x)^2}{\delta_{0h}^2}\phi(x - Z; 0, \delta_{0h}^2)$$

$$+\left[\frac{\lambda\sigma_u^2 h^2}{\lambda\sigma_u^2+h^2}+\left(\frac{\lambda\sigma_u^2(x-Z)}{\lambda\sigma_u^2+h^2}\right)^2\right]\phi(x-Z;0,\delta_{0h}^2),$$

this is exactly (3.3.5). □

For the sake of brevity, denote $\delta_{jh}=\sqrt{h^2+(\lambda+j)\sigma_u^2}$, $j=0,1,2,3$.

1. To calculate the asymptotic variance of $v_{i\lambda}(x)$ defined in (3.4.4), $i=1,2,\ldots,n$, and compare the difference between $\hat{g}_n(x;\lambda)$ and $\tilde{g}_n(x;\lambda)$.

$$\begin{aligned}
&E(Z-x)\phi^2(x,Z,\delta_{0h}^2)=\iint(t+u-x)\phi^2(x;t+u,\delta_{0h}^2)\phi(u;0,\sigma_u^2)f_X(t)dudt\\
&=\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\iint(t+u-x)\phi\left(u;x-t,\frac{\delta_{0h}^2}{2}\right)\phi(u;0,\sigma_u^2)f_X(t)dudt\\
&=\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\iint(t+u-x)\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)\cdot\phi\left(u;\frac{2\sigma_u^2(x-t)}{\delta_{2h}^2},\frac{\sigma_u^2\delta_{0h}^2}{\delta_{2h}^2}\right)f_X(t)dudt\\
&=\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\int(t-x)\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt-\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\cdot\frac{2\sigma_u^2}{\delta_{2h}^2}\int(t-x)\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt\\
&=\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\left(1-\frac{2\sigma_u^2}{\delta_{2h}^2}\right)\int(t-x)\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt\\
&=\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\left(\frac{\delta_{0h}^2}{\delta_{2h}^2}\right)\int(t-x)\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt\\
&=\frac{\delta_{0h}}{2\sqrt{\pi}\delta_{2h}^2}\left[f_{1,\frac{\lambda}{2}}(x)+\frac{h^2}{4}f''_{1,\frac{\lambda}{2}}(x)-xf_{0,\frac{\lambda}{2}}(x)-x\frac{h^2}{4}f''_{0,\frac{\lambda}{2}}(x)+o(h^2)\right].
\end{aligned}$$

Next, note that

$$\begin{aligned}
&E[Y\phi^2(x;Z,\delta_{0h}^2)]=E\left(E[(g(X)+\varepsilon)\phi^2(x;Z,\delta_{0h}^2)\Big|X,U]\right)=E[g(X)\phi^2(x;Z,\delta_{0h}^2)]\\
&=\iint g(t)\phi^2(x;t+u,\delta_{0h}^2)\phi(u;0,\sigma_u^2)f_X(t)dudt\\
&=\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\iint g(t)\phi\left(u;x-t,\frac{\delta_{0h}^2}{2}\right)\phi(u;0,\sigma_u^2)f_X(t)dudt\\
&=\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\iint g(t)\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)\phi\left(u;\frac{2\sigma_u^2(x-t)}{\delta_{2h}^2},\frac{\sigma_u^2\delta_{0h}^2}{\delta_{2h}^2}\right)f_X(t)dudt\\
&=\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\int g(t)\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt=\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\left[g_{0,\frac{\lambda}{2}}(x)+\frac{h^2}{4}g''_{0,\frac{\lambda}{2}}(x)+o(h^2)\right]
\end{aligned}$$

Now, we have

$$
\begin{aligned}
&E[Y(Z-x)\phi^2(x;Z,\delta_{0h}^2)] = E[(g(X)+\varepsilon)(Z-x)\phi^2(x;Z,\delta_{0h}^2)] \\
&= E\left((Z-x)\phi^2(x;Z,\delta_{0h}^2)E[(g(X)+\varepsilon)\Big|X,U]\right) = E[g(X)(Z-x)\phi^2(x;Z,\delta_{0h}^2)] \\
&= \iint g(t)(t+u-x)\phi^2(x;t+u,\delta_{0h}^2)\phi(u;0,\sigma_u^2)f_X(t)dudt \\
&= \frac{1}{2\sqrt{\pi\delta_{0h}^2}}\iint g(t)(t+u-x)\phi\left(u;x-t,\frac{\delta_{0h}^2}{2}\right)\phi(u;0,\sigma_u^2)f_X(t)dudt \\
&= \frac{1}{2\sqrt{\pi\delta_{0h}^2}}\iint g(t)(t+u-x)\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)\phi\left(u;\frac{2\sigma_u^2(x-t)}{\delta_{2h}^2},\frac{\sigma_u^2\delta_{0h}^2}{\delta_{2h}^2}\right)f_X(t)dudt \\
&= \frac{1}{2\sqrt{\pi\delta_{0h}^2}}\int g(t)(t-x)\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt - \frac{2\sigma_u^2}{\delta_{2h}^2\sqrt{2\pi\delta_{0h}^2}}\int g(t)(t-x)\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt \\
&= \frac{1}{2\sqrt{\pi\delta_{0h}^2}}\left(1-\frac{2\sigma_u^2}{\delta_{2h}^2}\right)\int g(t)(t-x)\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt \\
&= \frac{1}{2\sqrt{\pi\delta_{0h}^2}}\left(\frac{\delta_{0h}^2}{\delta_{2h}^2}\right)\int g(t)(t-x)\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt \\
&= \frac{\delta_{0h}}{2\sqrt{\pi}\delta_{2h}^2}\left[g_{1,\frac{\lambda}{2}}(x)+\frac{h^2}{4}g''_{1,\frac{\lambda}{2}}(x)-xg_{0,\frac{\lambda}{2}}(x)-x\frac{h^2}{4}g''_{0,\frac{\lambda}{2}}(x)+o(h^2)\right]
\end{aligned}
$$

For the fourth one, note that

$$
\begin{aligned}
&E[Y(Z-x)^2\phi^2(x;Z,\delta_{0h}^2)] = E[(g(X)+\varepsilon)(Z-x)^2\phi^2(x;Z,\delta_{0h}^2)] \\
&= E\left((Z-x)^2\phi^2(x;Z,\delta_{0h}^2)E[(g(X)+\varepsilon)\Big|X,U]\right) = E[g(X)(Z-x)^2\phi^2(x;Z,\delta_{0h}^2)] \\
&= \iint g(t)(t+u-x)^2\phi^2(x;t+u,\delta_{0h}^2)\phi(u;0,\sigma_u^2)f_X(t)dudt \\
&= \frac{1}{2\sqrt{\pi\delta_{0h}^2}}\iint g(t)(t+u-x)^2\phi\left(u;x-t,\frac{\delta_{0h}^2}{2}\right)\phi(u;0,\sigma_u^2)f_X(t)dudt \\
&= \frac{1}{2\sqrt{\pi\delta_{0h}^2}}\iint g(t)(t+u-x)^2\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)\phi\left(u;\frac{2\sigma_u^2(x-t)}{\delta_{2h}^2},\frac{\sigma_u^2\delta_{0h}^2}{\delta_{2h}^2}\right)f_X(t)dudt \\
&= \frac{1}{2\sqrt{\pi\delta_{0h}^2}}\int g(t)(t-x)^2\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt - \frac{\sigma_u^2}{\delta_{2h}^2\sqrt{\pi\delta_{0h}^2}}\int g(t)(t-x)^2\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt \\
&\quad + \frac{1}{2\sqrt{\pi\delta_{0h}^2}}\int g(t)\left(\frac{\sigma_u^2\delta_{0h}^2}{\delta_{2h}^2}+\left[\frac{2\sigma_u^2(x-t)}{\delta_{2h}^2}\right]^2\right)\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt \\
&= \frac{1}{2\sqrt{\pi\delta_{0h}^2}}\left(1-\frac{4\sigma_u^2}{\delta_{2h}^2}+\frac{4\sigma^4}{\delta_{2h}^4}\right)\int g(t)(t-x)^2\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt
\end{aligned}
$$

$$+\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\cdot\frac{\sigma_u^2\delta_{0h}^2}{\delta_{2h}^2}\int g(t)\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt$$

$$=\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\left(1-\frac{2\sigma_u^2}{\delta_{2h}^2}\right)^2\int g(t)(t-x)^2\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt$$

$$+\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\cdot\frac{\sigma_u^2\delta_{0h}^2}{\delta_{2h}^2}\int g(t)\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt$$

$$=\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\left(\frac{\delta_{0h}^2}{\delta_{2h}^2}\right)^2\int g(t)(t-x)^2\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt$$

$$+\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\cdot\frac{\sigma_u^2\delta_{0h}^2}{\delta_{2h}^2}\int g(t)\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt$$

$$=\frac{\delta_{0h}^3}{2\sqrt{\pi}\delta_{2h}^4}\left[g_{2,\frac{\lambda}{2}}(x)+\frac{h^2}{4}g''_{2,\frac{\lambda}{2}}(x)-xg_{1,\frac{\lambda}{2}}(x)-x\frac{h^2}{4}g''_{1,\frac{\lambda}{2}}(x)+x^2g_{0,\frac{\lambda}{2}}(x)+x^2\frac{h^2}{4}g''_{0,\frac{\lambda}{2}}(x)+o(h^2)\right]$$

$$+\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\cdot\frac{\sigma_u^2\delta_{0h}^2}{\delta_{2h}^2}\left[g_{0,\frac{\lambda}{2}}(x)+\frac{h^2}{4}g''_{0,\frac{\lambda}{2}}(x)+o(h^2)\right].$$

For the fifth one, we have

$$E[Y^2(Z-x)\phi^2(x;Z,\delta_{0h}^2)]=E[(g(X)+\varepsilon)^2(Z-x)\phi^2(x;Z,\delta_{0h}^2)]$$

$$=E[(g^2(X)+\tau^2(X))(Z-x)\phi^2(x;Z,\delta_{0h}^2)]$$

$$=\iint(g^2(t)+\tau^2(t))(t+u-x)\phi^2(x;t+u,\delta_{0h}^2)\phi(u;0,\sigma_u^2)f_X(t)dudt$$

$$=\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\iint(g^2(t)+\tau^2(t))(t+u-x)\phi\left(u;x-t,\frac{\delta_{0h}^2}{2}\right)\phi(u;0,\sigma_u^2)f_X(t)dudt$$

$$=\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\iint(g^2(t)+\tau^2(t))(t+u-x)\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)\phi\left(u;\frac{2\sigma_u^2(x-t)}{\delta_{2h}^2},\frac{\sigma_u^2\delta_{0h}^2}{\delta_{2h}^2}\right)f_X(t)dudt$$

$$=\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\int(g^2(t)+\tau^2(t))(t-x)\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt$$

$$-\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\cdot\frac{2\sigma_u^2}{\delta_{2h}^2}\int(g^2(t)+\tau^2(t))(t-x)\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt$$

$$=\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\left(1-\frac{2\sigma_u^2}{\delta_{2h}^2}\right)\int(g^2(t)+\tau^2(t))(t-x)\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt$$

$$=\frac{1}{2\sqrt{\pi\delta_{0h}^2}}\left(\frac{\delta_{0h}^2}{\delta_{2h}^2}\right)\int(g^2(t)+\tau^2(t))(t-x)\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt$$

$$=\frac{\delta_{0h}}{2\sqrt{\pi}\delta_{2h}^2}\int t(g^2(t)+\tau^2(t))\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt-\frac{\delta_{0h}x}{2\sqrt{\pi}\delta_{2h}^2}\int(g^2(t)+\tau^2(t))\phi\left(t,x,\frac{\delta_{2h}^2}{2}\right)f_X(t)dt$$

$$=\frac{\delta_{0h}}{2\sqrt{\pi}\delta_{2h}^2}\left[G_{1,\frac{\lambda}{2}}(x)+\frac{h^2}{4}G''_{1,\frac{\lambda}{2}}(x)-xG_{0,\frac{\lambda}{2}}(x)-\frac{xh^2}{4}G''_{0,\frac{\lambda}{2}}(x)+o(h^2)\right]$$

$$+\frac{\delta_{0h}}{2\sqrt{\pi}\delta_{2h}^2}\left[H_{1,\frac{\lambda}{2}}(x)+\frac{h^2}{4}H''_{1,\frac{\lambda}{2}}(x)-xH_{0,\frac{\lambda}{2}}(x)-\frac{xh^2}{4}H''_{0,\frac{\lambda}{2}}(x)+o(h^2)\right].$$

Recall that for each $i=1,2,\ldots,n$,

$$\begin{aligned}
\xi_{0\lambda,i}(x) &= \phi(x,Z_i,\delta_{0h}^2)-E\phi(x,Z,\delta_{0h}^2),\\
\xi_{1\lambda,i}(x) &= \frac{h^2}{\delta_{0h}^2}\left[(Z_i-x)\phi(x,Z_i,\delta_{0h}^2)-E(Z-x)\phi(x,Z,\delta_{0h}^2)\right],\\
\xi_{2\lambda,i}(x) &= \frac{h^4}{\delta_{0h}^4}\left[(Z_i-x)^2\phi(x,Z_i,\delta_{0h}^2)-E(Z-x)^2\phi(x,Z,\delta_{0h}^2)\right]\\
&\quad+\frac{\lambda\sigma_u^2h^2}{\delta_{0h}^2}\left[\phi(x,Z_i,\delta_{0h}^2)-E\phi(x,Z,\delta_{0h}^2)\right],\\
\eta_{0\lambda,i}(x) &= Y_i\phi(x,Z_i,\delta_{0h}^2)-EY\phi(x,Z,\delta_{0h}^2)\\
\eta_{1\lambda,i}(x) &= \frac{h^2}{\delta_{0h}^2}\left[Y_i(Z_i-x)\phi(x,Z_i,\delta_{0h}^2)-EY(Z-x)\phi(x,Z,\delta_{0h}^2)\right].
\end{aligned}$$

Then

$$\begin{aligned}
&\operatorname{Cov}\left(\xi_{0,\lambda_1}(x),\xi_{0,\lambda_2}(x)\right)=E\left(\xi_{0,\lambda_1}(x)\cdot\xi_{0,\lambda_2}(x)\right)-E\xi_{0,\lambda_1}(x)\cdot E\xi_{0,\lambda_2}(x)\\
&=E\left[\phi(x,Z,h^2+\lambda_1\sigma_u^2)\cdot Y\phi(x,Z,h^2+\lambda_2\sigma_u^2)\right]-E\phi(x,Z,h^2+\lambda_1\sigma_u^2)\cdot EY\phi(x,Z,h^2+\lambda_2\sigma_u^2)
\end{aligned}$$

For the sake of brevity, in the following, we shall denote

$$\alpha_{jh}^2=h^2+\lambda_j\sigma_u^2,\quad j=1,2,\quad \alpha_{12h}^2=\frac{\alpha_{1h}^2\alpha_{2h}^2}{\alpha_{1h}^2+\alpha_{2h}^2},\quad \lambda_{12}=\frac{\lambda_1+\lambda_2}{\lambda_1\lambda_2}.$$

Note that

$$\begin{aligned}
&E\left[\phi(x,Z,\alpha_{1h}^2)\cdot\phi(x,Z,\alpha_{2h}^2)\right]=\iint\phi(x,t+u,\alpha_{1h}^2)\phi(x,t+u,\alpha_{2h}^2)\phi(u,0,\sigma_u^2)f_X(t)dudt\\
&=\iint\phi(u,x-t,\alpha_{1h}^2)\phi(u,x-t,\alpha_{2h}^2)\phi(u,0,\sigma_u^2)f_X(t)dudt\\
&=\iint\phi\left(u,x-t,\alpha_{12h}^2\right)\phi\left(0,0,\alpha_{1h}^2+\alpha_{2h}^2\right)\phi(u,0,\sigma_u^2)f_X(t)dudt\\
&=\frac{1}{\sqrt{2\pi(\alpha_{1h}^2+\alpha_{2h}^2)}}\iint\phi\left(u,x-t,\alpha_{12h}^2\right)\phi(u,0,\sigma_u^2)f_X(t)dudt
\end{aligned}$$

$$= \frac{1}{\sqrt{2\pi(\alpha_{1h}^2+\alpha_{2h}^2)}} \iint \phi\left(u, \frac{(x-t)\sigma_u^2}{\alpha_{12h}^2+\sigma_u^2}, \frac{\alpha_{1h}^2\alpha_{2h}^2\sigma_u^2}{\alpha_{12h}^2+\sigma_u^2}\right) \cdot$$
$$\phi\left(t; x, \alpha_{12h}^2+\sigma_u^2\right) f_X(t)dudt$$
$$= \frac{1}{\sqrt{2\pi(\alpha_{1h}^2+\alpha_{2h}^2)}} \int \phi\left(t; x, \alpha_{12h}^2+\sigma_u^2\right) f_X(t)dt$$
$$\to \frac{1}{\sqrt{2\pi\left(\lambda_1+\lambda_2\right)\sigma_u^2}} \int \phi\left(t; x, \left(\lambda_{12}+1\right)\sigma_u^2\right) f_X(t)dt$$

as $h \to 0$, so we have

$$\mathrm{Cov}(\xi_{0,\lambda_1}, \xi_{0,\lambda_2}) = \frac{1}{\sqrt{2\pi\left(\lambda_1+\lambda_2\right)\sigma_u^2}} \int \phi\left(t, x, \left(\lambda_{12}+1\right)\sigma_u^2\right) f_X(t)dt - f_{0,\lambda_1}(x)f_{0,\lambda_2}(x) + o(1).$$

Next,

$$\mathrm{Cov}\left(\xi_{0,\lambda_1}(x), \eta_{0,\lambda_2}(x)\right) = E\left(\xi_{0,\lambda_1}(x)\cdot\eta_{0,\lambda_2}(x)\right) - E\xi_{0,\lambda_1}(x)\cdot E\eta_{0,\lambda_2}(x)$$
$$= E\left[\phi(x, Z, \alpha_{1h}^2)\cdot Y\phi(x, Z, \alpha_{2h}^2)\right] - E\phi(x, Z, \alpha_{1h}^2)\cdot EY\phi(x, Z, \alpha_{2h}^2),$$

we have

$$E\left[\phi(x, Z_i, \alpha_{1h}^2)\cdot Y\phi(x, Z_i, \alpha_{2h}^2)\right] = \iint g(t)\phi(x, t+u, \alpha_{1h}^2)\phi(x, t+u, \alpha_{2h}^2)\phi(u, 0, \sigma_u^2)f_X(t)dudt$$
$$= \iint g(t)\phi(u, x-t, \alpha_{1h}^2)\phi(u, x-t, \alpha_{2h}^2)\phi(u, 0, \sigma_u^2)f_X(t)dudt$$
$$= \iint g(t)\phi\left(u, x-t, \alpha_{12h}^2\right)\phi\left(0, 0, \alpha_{1h}^2+\alpha_{2h}^2\right)\phi(u, 0, \sigma_u^2)f_X(t)dudt$$
$$= \frac{1}{\sqrt{2\pi(\alpha_{1h}^2+\alpha_{2h}^2)}} \iint g(t)\phi\left(u, x-t, \alpha_{12h}^2\right)\phi(u, 0, \sigma_u^2)f_X(t)dudt$$
$$= \frac{1}{\sqrt{2\pi(\alpha_{1h}^2+\alpha_{2h}^2)}} \iint g(t)\phi\left(u, \frac{(x-t)\sigma_u^2}{\alpha_{12h}^2+\sigma_u^2}, \frac{\alpha_{1h}^2\alpha_{2h}^2\sigma_u^2}{\alpha_{12h}^2+\sigma_u^2}\right) \cdot$$
$$\phi\left(t; x, \alpha_{12h}^2+\sigma_u^2\right) f_X(t)dudt$$
$$= \frac{1}{\sqrt{2\pi(\alpha_{1h}^2+\alpha_{2h}^2)}} \int g(t)\phi\left(t, x, \alpha_{12h}^2+\sigma_u^2\right) f_X(t)dt$$
$$\to \frac{1}{\sqrt{2\pi\left(\lambda_1+\lambda_2\right)\sigma_u^2}} \int g(t)\phi\left(t, x, \left(\lambda_{12}+1\right)\sigma_u^2\right) f_X(t)dt$$

as $h \to 0$. Hence,

$$\mathrm{Cov}(\xi_{0,\lambda_1}, \eta_{0,\lambda_2}) = \frac{1}{\sqrt{2\pi\left(\lambda_1+\lambda_2\right)\sigma_u^2}} \int g(t)\phi\left(t, x, \sigma_u^2(\lambda_{12}+1)\right) f_X(t)dt - f_{0,\lambda_1}(x) g_{0,\lambda_2}(x) + o(1).$$

Next, we have

$$\begin{aligned} &\mathrm{Cov}\left(\eta_{0,\lambda_1}(x), \eta_{0,\lambda_2}(x)\right) = E\left(\eta_{0,\lambda_1}(x)\cdot\eta_{0,\lambda_2}(x)\right) - E\eta_{0,\lambda_1}(x)\cdot E\eta_{0,\lambda_2}(x) \\ &= E\left[Y\phi(x,Z,\alpha_{1h}^2)\cdot Y\phi(x,Z,\alpha_{2h}^2)\right] - EY\phi(x,Z,\alpha_{1h}^2)\cdot EY\phi(x,Z,\alpha_{2h}^2), \end{aligned}$$

and

$$\begin{aligned} &E\left[Y\phi(x,Z,\alpha_{1h}^2)\cdot Y\phi(x,Z,\alpha_{2h}^2)\right] \\ &= \iint g^2(t)\phi(x,t+u,\alpha_{1h}^2)\phi(x,t+u,\alpha_{2h}^2)\phi(u,0,\sigma_u^2)f_X(t)dudt \\ &= \iint g^2(t)\phi(u,x-t,\alpha_{1h}^2)\phi(u,x-t,\alpha_{2h}^2)\phi(u,0,\sigma_u^2)f_X(t)dudt \\ &= \iint g^2(t)\phi\left(u,x-t,\alpha_{12h}^2\right)\phi\left(0,0,\alpha_{1h}^2+\alpha_{2h}^2\right)\phi(u,0,\sigma_u^2)f_X(t)dudt \\ &= \frac{1}{\sqrt{2\pi(\alpha_{1h}^2+\alpha_{2h}^2)}}\iint g^2(t)\phi\left(u,x-t,\alpha_{12h}^2\right)\phi(u,0,\sigma_u^2)f_X(t)dudt \\ &= \frac{1}{\sqrt{2\pi(\alpha_{1h}^2+\alpha_{2h}^2)}}\iint \phi\left(u,\frac{(x-t)\sigma_u^2}{\alpha_{12h}^2+\sigma_u^2},\frac{\alpha_{1h}^2\alpha_{2h}^2\sigma_u^2}{\alpha_{12h}^2+\sigma_u^2}\right)\cdot\phi\left(t,x,\alpha_{12h}^2+\sigma_u^2\right)g^2(t)f_X(t)dudt \\ &= \frac{1}{\sqrt{2\pi(\alpha_{1h}^2+\alpha_{2h}^2)}}\int g^2(t)\phi\left(t,x,\alpha_{12h}^2+\sigma_u^2\right)f_X(t)dt \\ &\to \frac{1}{\sqrt{2\pi\left(\lambda_1+\lambda_2\right)\sigma_u^2}}\int g^2(t)\phi\left(t,x,\left(\lambda_{12}+1\right)\sigma_u^2\right)f_X(t)dt \end{aligned}$$

as $h \to 0$. Hence

$$\mathrm{Cov}(\eta_{0,\lambda_1}, \eta_{0,\lambda_2}) = \frac{1}{\sqrt{2\pi\left(\lambda_1+\lambda_2\right)\sigma_u^2}} \int g^2(t)\phi\left(t, x, (\lambda_{12}+1)\sigma_u^2\right) f_X(t)dt - g_{0,\lambda_1}(x) g_{0,\lambda_2}(x) + o(1).$$

Next, we see

$$\mathrm{Cov}\left(\xi_{0,\lambda_1}(x), \xi_{1,\lambda_2}(x)\right) = E\left(\xi_{0,\lambda_1}(x)\cdot\xi_{1,\lambda_2}(x)\right) - E\xi_{0,\lambda_1}(x)\cdot E\xi_{1,\lambda_2}(x)$$

$$= E\left[\phi(x,Z,\alpha_{1h}^2)\cdot\frac{h^2}{\alpha_{2h}^2}(Z-x)\phi(x,Z,\alpha_{2h}^2)\right]-\frac{h^2}{\alpha_{2h}^2}\cdot E\phi(x,Z,\alpha_{1h}^2)E(Z-x)\phi(x,Z,\alpha_{2h}^2).$$

Note that when $\lambda_2\neq 0$, $\operatorname{Cov}(\xi_{0,\lambda_1}(x),\xi_{1,\lambda_2}(x))=O(h^2)$. When $\lambda_2=0$, we have

$$\begin{aligned}
&E\left[\phi(x,Z,\alpha_{1h}^2)\cdot(Z-x)\phi(x,Z,h^2)\right]\\
&=\iint(t+u-x)\phi(u,x-t,\alpha_{1h}^2)\phi(u,x-t,h^2)\phi(u,0,\sigma_u^2)f_X(t)dudt\\
&=\frac{1}{\sqrt{2\pi(h^2+\alpha_{1h}^2)}}\iint(t+u-x)\phi\left(u,x-t,\frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}\right)\phi(u,0,\sigma_u^2)f_X(t)dudt\\
&=\frac{1}{\sqrt{2\pi(h^2+\alpha_{1h}^2)}}\iint(t+u-x)\phi\left(t,x,\frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2\right)\cdot\\
&\qquad\phi\left(u,\frac{(x-t)\sigma_u^2}{\frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2},\frac{\alpha_{1h}^2h^2\sigma_u^2}{\frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2}\right)f_X(t)dudt\\
&=\frac{1}{\sqrt{2\pi(h^2+\alpha_{1h}^2)}}\int(t-x)\left[1-\frac{\sigma_u^2}{\frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2}\right]\phi\left(t,x,\frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2\right)f_X(t)dt\\
&=\frac{h^2\alpha_{1h}^2}{h^2\alpha_{1h}^2+(h^2+\alpha_{1h}^2)\sigma_u^2}\cdot\frac{1}{\sqrt{2\pi(h^2+\alpha_{1h}^2)}}\int(t-x)\phi\left(t,x,\frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2\right)f_X(t)dt,
\end{aligned}$$

which is $O(h^2)$. Therefore, for all λ_1 and λ_2, $\operatorname{Cov}(\xi_{0,\lambda_1},\xi_{1,\lambda_2})=O(h^2)$. Next, we see

$$\begin{aligned}
&\operatorname{Cov}(\xi_{0,\lambda_1}(x),\eta_{1,\lambda_2}(x))=E(\xi_{0,\lambda_1}(x)\cdot\eta_{1,\lambda_2}(x))-E\xi_{0,\lambda_1}(x)\cdot E\eta_{1,\lambda_2}(x)\\
&=E\left[\phi(x,Z,\alpha_{1h}^2)\cdot\frac{h^2}{\alpha_{2h}^2}Y(Z-x)\phi(x,Z,\alpha_{2h}^2)\right]-\frac{h^2}{\alpha_{2h}^2}\cdot E\phi(x,Z,\alpha_{1h}^2)EY(Z-x)\phi(x,Z,\alpha_{2h}^2).
\end{aligned}$$

Note that when $\lambda_2\neq 0$, $\operatorname{Cov}(\xi_{0,\lambda_1}(x),\eta_{1,\lambda_2}(x))=O(h^2)$. When $\lambda_2=0$, we have

$$\begin{aligned}
&E\left[\phi(x,Z,\alpha_{1h}^2)\cdot Y(Z-x)\phi(x,Z,h^2)\right]\\
&=\iint g(t)(t+u-x)\phi(u,x-t,\alpha_{1h}^2)\phi(u,x-t,h^2)\phi(u,0,\sigma_u^2)f_X(t)dudt\\
&=\frac{1}{\sqrt{2\pi(h^2+\alpha_{1h}^2)}}\iint g(t)(t+u-x)\phi\left(u,x-t,\frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}\right)\phi(u,0,\sigma_u^2)f_X(t)dudt\\
&=\frac{1}{\sqrt{2\pi(h^2+\alpha_{1h}^2)}}\iint g(t)(t+u-x)\phi\left(t,x,\frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2\right)\cdot
\end{aligned}$$

$$\phi\left(u, \frac{(x-t)\sigma_u^2}{\frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2}, \frac{\alpha_{1h}^2 h^2\sigma_u^2}{\frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2}\right) f_X(t)dudt$$

$$= \frac{1}{\sqrt{2\pi(h^2+\alpha_{1h}^2)}}\int g(t)(t-x)\left[1-\frac{\sigma_u^2}{\frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2}\right]\phi\left(t;x,\frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2\right) f_X(t)dt$$

$$= \frac{h^2\alpha_{1h}^2}{(h^2\alpha_{1h}^2+\sigma_u^2(h^2+\alpha_{1h}^2))\sqrt{2\pi(h^2+\alpha_{1h}^2)}}\int \phi\left(t,x,\frac{h^2\alpha_{1h}^2}{2\alpha_{1h}^2}+\sigma_u^2\right) g(t)(t-x)f_X(t)dt,$$

which is $O(h^2)$. So, $\text{Cov}(\xi_{0,\lambda_1}, \eta_{1,\lambda_2}) = O(h^2)$ for all λ_1, λ_2. Next, note that

$$\text{Cov}\left(\xi_{1,\lambda_1}(x), \eta_{0,\lambda_2}(x)\right) = E\left(\xi_{1,\lambda_1}(x)\cdot\eta_{0,\lambda_2}(x)\right) - E\xi_{1,\lambda_1}(x)\cdot E\eta_{0,\lambda_2}(x)$$
$$= E\left[\frac{h^2}{\alpha_{1h}^2}(Z-x)\phi(x,Z,\alpha_{1h}^2)\cdot Y\phi(x,Z,\alpha_{2h}^2)\right] - \frac{h^2}{\alpha_{1h}^2}E(Z-x)\phi(x,Z,\alpha_{1h}^2)\cdot EY\phi(x,Z,\alpha_{2h}^2).$$

Note that when $\lambda_1 \neq 0$, $\text{Cov}\left(\xi_{0,\lambda_1}(x), \eta_{1,\lambda_2}(x)\right) = O(h^2)$. When $\lambda_1 = 0$, we have

$$E\left[(Z-x)\phi(x,Z,h^2)\cdot Y\phi(x,Z,\alpha_{2h}^2)\right]$$
$$= \iint g(t)(t+u-x)\phi(u,x-t,h^2)\phi(u,x-t,\alpha_{2h}^2)\phi(u,0,\sigma_u^2)f_X(t)dudt$$
$$= \frac{1}{\sqrt{2\pi(h^2+\alpha_{2h}^2)}}\iint g(t)(t+u-x)\phi\left(u,x-t,\frac{h^2\alpha_{2h}^2}{h^2+\alpha_{2h}^2}\right)\phi(u,0,\sigma_u^2)f_X(t)dudt$$
$$= \frac{1}{\sqrt{2\pi(h^2+\alpha_{2h}^2)}}\iint g(t)(t+u-x)\phi\left(t;x,\frac{h^2\alpha_{2h}^2}{h^2+\alpha_{2h}^2}+\sigma_u^2\right)\cdot$$
$$\phi\left(u, \frac{(x-t)\sigma_u^2}{\frac{h^2\alpha_{2h}^2}{h^2+\alpha_{2h}^2}+\sigma_u^2}, \frac{\alpha_{2h}^2 h^2\sigma_u^2}{\frac{h^2\alpha_{2h}^2}{h^2+\alpha_{2h}^2}+\sigma_u^2}\right) f_X(t)dudt$$
$$= \frac{1}{\sqrt{2\pi(h^2+\alpha_{2h}^2)}}\int g(t)(t-x)\left[1-\frac{\sigma_u^2}{\frac{h^2\alpha_{2h}^2}{h^2+\alpha_{2h}^2}+\sigma_u^2}\right]\phi\left(t,x,\frac{h^2\alpha_{2h}^2}{h^2+\alpha_{2h}^2}+\sigma_u^2\right) f_X(t)dt$$
$$= \frac{h^2\alpha_{2h}^2}{(h^2\alpha_{2h}^2+\sigma_u^2(h^2+\alpha_{2h}^2))\sqrt{2\pi(h^2+\alpha_{2h}^2)}}\int \phi\left(t,x,\frac{h^2\alpha_{2h}^2}{h^2+\alpha_{2h}^2}+\sigma_u^2\right) g(t)(t-x)f_X(t)dt$$

which is the order of $O(h^2)$. Therefore, $\text{Cov}(\xi_{1,\lambda_1}, \eta_{0,\lambda_2}) = O(h^2)$ for all λ_1, λ_2. Next, note that

$$\text{Cov}\left(\eta_{1,\lambda_1}(x), \xi_{0,\lambda_2}(x)\right) = E\left(\eta_{1,\lambda_1}(x)\cdot\xi_{0,\lambda_2}(x)\right) - E\eta_{1,\lambda_1}(x)\cdot E\xi_{0,\lambda_2}(x)$$

$$= E\left[\frac{h^2}{\alpha_{1h}^2} Y(Z-x)\phi(x, Z_i, \alpha_{1h}^2)\cdot\phi(x, Z, \alpha_{2h}^2)\right] - \frac{h^2}{\alpha_{1h}^2} EY(Z-x)\phi(x, Z, \alpha_{1h}^2)\cdot E\phi(x, Z, \alpha_{2h}^2).$$

Note that when $\lambda_1 \neq 0$, $\operatorname{Cov}(\eta_{1,\lambda_1}(x), \xi_{0,\lambda_2}(x)) = O(h^2)$. When $\lambda_1 = 0$, we have

$$\begin{aligned}
& E\left[Y(Z-x)\phi(x, Z, h^2)\cdot\phi(x, Z, \alpha_{2h}^2)\right] \\
&= \iint g(t)(t+u-x)\phi(u, x-t, h^2)\phi(u, x-t, \alpha_{2h}^2)\phi(u, 0, \sigma_u^2) f_X(t)\,dudt \\
&= \frac{1}{\sqrt{2\pi(h^2+\alpha_{2h}^2)}} \iint g(t)(t+u-x)\phi\left(u, x-t, \frac{h^2\alpha_{2h}^2}{h^2+\alpha_{2h}^2}\right)\phi(u, 0, \sigma_u^2) f_X(t)\,dudt \\
&= \frac{1}{\sqrt{2\pi(h^2+\alpha_{2h}^2)}} \iint g(t)(t+u-x)\phi\left(t; x, \frac{h^2\alpha_{2h}^2}{h^2+\alpha_{2h}^2}+\sigma_u^2\right)\cdot \\
&\qquad \phi\left(u, \frac{(x-t)\sigma_u^2}{\frac{h^2\alpha_{2h}^2}{h^2+\alpha_{2h}^2}+\sigma_u^2}, \frac{\alpha_{2h}^2 h^2\sigma_u^2}{\frac{h^2\alpha_{2h}^2}{h^2+\alpha_{2h}^2}+\sigma_u^2}\right) f_X(t)\,dudt \\
&= \frac{1}{\sqrt{2\pi(h^2+\alpha_{2h}^2)}} \int g(t)(t-x)\left[1 - \frac{\sigma_u^2}{\frac{h^2\alpha_{2h}^2}{h^2+\alpha_{2h}^2}+\sigma_u^2}\right]\phi\left(t, x, \frac{h^2\alpha_{2h}^2}{h^2+\alpha_{2h}^2}+\sigma_u^2\right) f_X(t)\,dt \\
&= \frac{h^2\alpha_{2h}^2}{(h^2\alpha_{2h}^2+\sigma_u^2(h^2+\alpha_{2h}^2)))\sqrt{2\pi(h^2+\alpha_{2h}^2)}} \int \phi\left(t, x, \frac{h^2\alpha_{2h}^2}{h^2+\alpha_{2h}^2}+\sigma_u^2\right)\cdot g(t)(t-x) f_X(t)\,dt,
\end{aligned}$$

which is $O(h^2)$. Therefore, $\operatorname{Cov}(\eta_{1,\lambda_1}, \xi_{0,\lambda_2}) = O(h^2)$ for all λ_1, λ_2.

Next, note that

$$\begin{aligned}
& \operatorname{Cov}(\eta_{0,\lambda_1}(x), \xi_{1,\lambda_2}(x)) = E(\eta_{0,\lambda_1}(x)\cdot\xi_{1,\lambda_2}(x)) - E\eta_{0,\lambda_1}(x)\cdot E\xi_{1,\lambda_2}(x) \\
&= E\left[Y\phi(x, Z, \alpha_{1h}^2)\cdot\frac{h^2}{\alpha_{2h}^2}(Z-x)\phi(x, Z, \alpha_{2h}^2)\right] - EY\phi(x, Z, \alpha_{1h}^2)\cdot\frac{h^2}{\alpha_{2h}^2}E(Z-x)\phi(x, Z, \alpha_{2h}^2).
\end{aligned}$$

When $\lambda_2 \neq 0$, $\operatorname{Cov}(\eta_{1,\lambda_1}(x), \xi_{0,\lambda_2}(x)) = O(h^2)$. When $\lambda_2 = 0$, we have

$$\begin{aligned}
& E\left[Y\phi(x, Z, \alpha_{1h}^2)\cdot(Z-x)\phi(x, Z, h^2)\right] \\
&= \iint g(t)(t+u-x)\phi(u, x-t, h^2)\phi(u, x-t, \alpha_{1h}^2)\phi(u, 0, \sigma_u^2) f_X(t)\,dudt \\
&= \frac{1}{\sqrt{2\pi(h^2+\alpha_{1h}^2)}} \iint g(t)(t+u-x)\phi\left(u, x-t, \frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}\right)\phi(u, 0, \sigma_u^2) f_X(t)\,dudt \\
&= \frac{1}{\sqrt{2\pi(h^2+\alpha_{1h}^2)}} \iint g(t)(t+u-x)\phi\left(t; x, \frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2\right)\cdot
\end{aligned}$$

$$\phi\left(u, \frac{(x-t)\sigma_u^2}{\frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2}, \frac{\alpha_{1h}^2 h^2\sigma_u^2}{\frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2}\right) f_X(t)dudt$$

$$= \frac{h^2\alpha_{1h}^2}{\left(h^2\alpha_{1h}^2+\sigma_u^2(h^2+\alpha_{1h}^2)\right)\sqrt{2\pi(h^2+\alpha_{1h}^2)}} \int \phi\left(t, x, \frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2\right) g(t)(t-x) f_X(t)dt,$$

which is $O(h^2)$. Therefore, we have $\mathrm{Cov}(\eta_{0,\lambda_1}, \xi_{1,\lambda_2}) = O(h^2)$ for all λ_1, λ_2. Next, we see that

$$\mathrm{Cov}\left(\eta_{0,\lambda_1}(x), \eta_{1,\lambda_2}(x)\right) = E\left(\eta_{0,\lambda_1}(x)\cdot\eta_{1,\lambda_2}(x)\right) - E\eta_{0,\lambda_1}(x)\cdot E\eta_{1,\lambda_2}(x)$$
$$= E\left[Y\phi(x,Z,\alpha_{1h}^2)\cdot\frac{h^2}{\alpha_{2h}^2}Y(Z-x)\phi(x,Z,\alpha_{2h}^2)\right] - EY\phi(x,Z,\alpha_{1h}^2)\cdot\frac{h^2}{\alpha_{2h}^2}EY(Z-x)\phi(x,Z,\alpha_{2h}^2).$$

Note that when $\lambda_2 \neq 0$, $\mathrm{Cov}\left(\eta_{1,\lambda_1}(x), \xi_{0,\lambda_2}(x)\right) = O(h^2)$. When $\lambda_2 = 0$, we have

$$E\left[Y\phi(x,Z,\alpha_{1h}^2)\cdot Y(Z-x)\phi(x,Z,h^2)\right]$$
$$= \iint g^2(t)(t+u-x)\phi(u,x-t,h^2)\phi(u,x-t,\alpha_{1h}^2)\phi(u,0,\sigma_u^2)f_X(t)dudt$$
$$= \frac{1}{\sqrt{2\pi(h^2+\alpha_{1h}^2)}}\iint g^2(t)(t+u-x)\phi\left(u, x-t, \frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}\right)\phi(u,0,\sigma_u^2)f_X(t)dudt$$
$$= \frac{1}{\sqrt{2\pi(h^2+\alpha_{1h}^2)}}\iint g^2(t)(t+u-x)\phi\left(t; x, \frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2\right)\cdot$$
$$\phi\left(u, \frac{(x-t)\sigma_u^2}{\frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2}, \frac{\alpha_{1h}^2 h^2\sigma_u^2}{\frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2}\right) f_X(t)dudt$$
$$= \frac{1}{\sqrt{2\pi(h^2+\alpha_{1h}^2)}}\int g^2(t)(t-x)\left[1-\frac{\sigma_u^2}{\frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2}\right]\phi\left(t, x, \frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2\right) f_X(t)dt$$
$$= \frac{h^2+\alpha_{1h}^2}{\left(h^2\alpha_{1h}^2+\sigma_u^2(h^2+\alpha_{1h}^2)\right)\sqrt{2\pi(h^2+\alpha_{1h}^2)}}\int \phi\left(t, x, \frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2\right)\cdot g^2(t)(t-x)f_X(t)dt,$$

which is the order of $O(h^2)$. Since, when $\lambda_2 = 0, EY(Z-x)\phi(x,Z,\alpha_{2h}^2) = O(h^2)$, we have $\mathrm{Cov}(\eta_{0,\lambda_1}, \eta_{1,\lambda_2}) = O(h^2)$ for all λ_1, λ_2. Next, we see that

$$\mathrm{Cov}\left(\xi_{0,\lambda_1}(x), \xi_{2,\lambda_2}(x)\right) = E\left(\xi_{0,\lambda_1}(x)\cdot\xi_{2,\lambda_2}(x)\right) - E\xi_{0,\lambda_1}(x)\cdot E\xi_{2,\lambda_2}(x)$$
$$= E\left[\phi(x,Z,\alpha_{1h}^2)\cdot\frac{h^4}{\alpha_{2h}^4}(Z-x)^2\phi(x,Z,\alpha_{2h}^2)\right] + E\left[\phi(x,Z,\alpha_{1h}^2)\cdot\frac{\lambda_2\sigma_u^2 h^2}{\alpha_{2h}^2}\phi(x,Z,\alpha_{2h}^2)\right]$$

$$-E\phi(x,Z,\alpha_{1h}^2)\cdot\frac{h^4}{\alpha_{2h}^4}E(Z-x)^2\phi(x,Z,\alpha_{2h}^2)-E\phi(x,Z,\alpha_{1h}^2)\cdot\frac{\lambda_2\sigma_u^2h^2}{\alpha_{2h}^2}E\phi(x,Z,\alpha_{2h}^2).$$

Note that when $\lambda_2\neq 0$, $\mathrm{Cov}\left(\xi_{0,\lambda_1}(x),\xi_{2,\lambda_2}(x)\right)=O(h^2)$. When $\lambda_2=0$, we have

$$\begin{aligned}
&E\left[\phi(x,Z,\alpha_{1h}^2)\cdot(Z-x)^2\phi(x,Z,h^2)\right]\\
&=\iint(t+u-x)^2\phi(u,x-t,h^2)\phi(u,x-t,\alpha_{1h}^2)\phi(u,0,\sigma_u^2)f_X(t)dudt\\
&=\frac{1}{\sqrt{2\pi(h^2+\alpha_{1h}^2)}}\iint(t+u-x)^2\phi\left(u,x-t,\frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}\right)\phi(u,0,\sigma_u^2)f_X(t)dudt\\
&=\frac{1}{\sqrt{2\pi(h^2+\alpha_{1h}^2)}}\iint(t+u-x)^2\phi\left(t;x,\frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2\right)\cdot\\
&\qquad\qquad\phi\left(u,\frac{(x-t)\sigma_u^2}{\frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2},\frac{\alpha_{1h}^2h^2\sigma_u^2}{\frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2}\right)f_X(t)dudt\\
&=\frac{1}{\sqrt{2\pi(h^2+\alpha_{1h}^2)}}\int(t-x)^2\left[1-\frac{\sigma_u^2}{\frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2}\right]^2\phi\left(t;x,\frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2\right)f_X(t)dt\\
&\quad+\frac{1}{\sqrt{2\pi(h^2+\alpha_{1h}^2)}}\frac{\sigma_u^2h^2\alpha_{1h}^2}{h^2\alpha_{1h}^2+\sigma_u^2(h^2+\alpha_{1h}^2)}\int\phi\left(t;x,\frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2\right)f_X(t)dt\\
&=\frac{(h^2\alpha_{1h}^2)^2}{(h^2\alpha_{1h}^2+\sigma_u^2(h^2+\alpha_{1h}^2))^2\sqrt{2\pi(h^2+\alpha_{1h}^2)}}\int(t-x)^2f_X(t)\phi\left(t,x,\frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2\right)dt\\
&\quad+\frac{1}{\sqrt{2\pi(h^2+\alpha_{1h}^2)}}\frac{\sigma_u^2h^2\alpha_{1h}^2}{h^2\alpha_{1h}^2+\sigma_u^2(h^2+\alpha_{1h}^2)}\int\phi\left(t,x,\frac{h^2\alpha_{1h}^2}{h^2+\alpha_{1h}^2}+\sigma_u^2\right)f_X(t)dt\\
&=O(h^2).
\end{aligned}$$

Since, when $\lambda_2=0, E(Z-x)^2\phi(x,Z,\alpha_{2h}^2)=O(h^2)$, we have $\mathrm{Cov}(\xi_{0,\lambda_1},\xi_{2,\lambda_2})=O(h^2)$ for all λ_1,λ_2. Next, note

$$\begin{aligned}
&\mathrm{Cov}\left(\xi_{2,\lambda_1}(x),\eta_{0,\lambda_2}(x)\right)=E\left(\xi_{2,\lambda_1}(x)\cdot\eta_{0,\lambda_2}(x)\right)-E\xi_{2,\lambda_1}(x)\cdot E\eta_{0,\lambda_2}(x)\\
&=E\left[\frac{h^4}{\alpha_{1h}^4}(Z-x)^2\phi(x,Z,\alpha_{1h}^2)\cdot Y\phi(x,Z,\alpha_{2h}^2)\right]+E\left[\frac{\lambda_1\sigma_u^2h^2}{\alpha_{1h}^2}\phi(x,Z,\alpha_{1h}^2)\cdot Y\phi(x,Z,\alpha_{2h}^2)\right]\\
&\quad-\frac{h^4}{\alpha_{1h}^4}E(Z-x)^2\phi(x,Z,\alpha_{1h}^2)\cdot EY\phi(x,Z,\alpha_{2h}^2)-\frac{\lambda_1\sigma_u^2h^2}{\alpha_{2h}^2}E\phi(x,Z,\alpha_{1h}^2)\cdot EY\phi(x,Z,\alpha_{2h}^2).
\end{aligned}$$

Note that when $\lambda_1 \neq 0$, $\text{Cov}\left(\xi_{2,\lambda_1}(x), \eta_{0,\lambda_2}(x)\right) = O(h^2)$. When $\lambda_1 = 0$, we have

$$\begin{aligned}
& E\left[(Z-x)^2\phi(x,Z,h^2)\cdot Y\phi(x,Z,\alpha_{2h}^2)\right] \\
=& \iint g(t)(t+u-x)^2\phi(u,x-t,h^2)\phi(u,x-t,\alpha_{2h}^2)\phi(u,0,\sigma_u^2)f_X(t)dudt \\
=& \frac{1}{\sqrt{2\pi(h^2+\alpha_{2h}^2)}}\iint g(t)(t+u-x)^2\phi\left(u,x-t,\frac{h^2\alpha_{2h}^2}{h^2+\alpha_{2h}^2}\right)\phi(u,0,\sigma_u^2)f_X(t)dudt \\
=& \frac{1}{\sqrt{2\pi(h^2+\alpha_{2h}^2)}}\iint g(t)(t+u-x)^2\phi\left(t;x,\frac{h^2\alpha_{2h}^2}{h^2+\alpha_{2h}^2}+\sigma_u^2\right)\cdot \\
& \qquad \phi\left(u,\frac{(x-t)\sigma_u^2}{\frac{h^2\alpha_{2h}^2}{h^2+\alpha_{2h}^2}+\sigma_u^2},\frac{\alpha_{2h}^2h^2\sigma_u^2}{\frac{h^2\alpha_{2h}^2}{h^2+\alpha_{2h}^2}+\sigma_u^2}\right)f_X(t)dudt \\
=& \frac{1}{\sqrt{2\pi(h^2+\alpha_{2h}^2)}}\int g(t)(t-x)^2\left[1-\frac{\sigma_u^2}{\frac{h^2\alpha_{2h}^2}{h^2+\alpha_{2h}^2}+\sigma_u^2}\right]^2\phi\left(t;x,\frac{h^2\alpha_{2h}^2}{h^2+\alpha_{2h}^2}+\sigma_u^2\right)f_X(t)dt \\
& +\frac{1}{\sqrt{2\pi(h^2+\alpha_{2h}^2)}}\frac{\sigma_u^2h^2\alpha_{2h}^2}{h^2\alpha_{2h}^2+\sigma_u^2(h^2+\alpha_{2h}^2)}\int g(t)\phi\left(t,x,\frac{h^2\alpha_{2h}^2}{h^2+\alpha_{2h}^2}+\sigma_u^2\right)f_X(t)dt \\
=& \frac{(h^2\alpha_{2h}^2)^2}{\left(h^2\alpha_{2h}^2+\sigma_u^2(h^2+\alpha_{2h}^2)\right)^2}\frac{1}{\sqrt{2\pi(h^2+\alpha_{2h}^2)}}\int(t-x)^2f_X(t)\phi\left(t,x,\frac{h^2\alpha_{2h}^2}{h^2+\alpha_{2h}^2}+\sigma_u^2\right)dt \\
& +\frac{1}{\sqrt{2\pi(h^2+\alpha_{2h}^2)}}\frac{\sigma_u^2h^2\alpha_{2h}^2}{h^2\alpha_{2h}^2+\sigma_u^2(h^2+\alpha_{2h}^2)}\int\phi\left(t,x,\frac{h^2\alpha_{2h}^2}{h^2+\alpha_{2h}^2}+\sigma_u^2\right)g(t)f_X(t)dt \\
=& O(h^2).
\end{aligned}$$

Therefore, we have for all λ_1, λ_2, $\text{Cov}(\xi_{2,\lambda_1}, \eta_{0,\lambda_2}) = O(h^2)$. Next, we look at

$$\begin{aligned}
& \text{Cov}(\xi_{1,\lambda_1},\xi_{1,\lambda_2}) = E(\xi_{1,\lambda_1}\cdot\xi_{1,\lambda_2}) - E(\xi_{1,\lambda_1})\cdot E(\xi_{1,\lambda_2}) \\
=& \frac{h^4}{\alpha_{1h}^2\alpha_{2h}^2}\left(E\left[(Z-x)^2\phi(x,Z,\alpha_{1h}^2)\phi(x,Z,\alpha_{2h}^2)\right]-\left[E(Z-x)\phi(x,Z,\alpha_{1h}^2)\cdot E(Z-x)\phi(x,Z,\alpha_{2h}^2)\right]\right).
\end{aligned}$$

Note that when $\lambda_1 = 0$ or $\lambda_2 = 0$, then $\text{Cov}(\xi_{1,\lambda_1}, \xi_{1,\lambda_2}) = O(h^2)$. When $\lambda_1 \neq 0$ and $\lambda_2 \neq 0$, then $\text{Cov}(\xi_{1,\lambda_1}, \xi_{1,\lambda_2}) = O(h^4)$. Similarly, we can show that when $\lambda_1 = 0$ or $\lambda_2 = 0$, then $\text{Cov}(\xi_{1,\lambda_1}, \eta_{1,\lambda_2})$, $\text{Cov}(\eta_{1,\lambda_1}, \eta_{1,\lambda_2})$, $\text{Cov}\left(\xi_{1,\lambda_1}(x), \xi_{2,\lambda_2}(x)\right)$, $\text{Cov}(\eta_{1,\lambda_1}, \xi_{2,\lambda_2})$, and $\text{Cov}\left(\xi_{2,\lambda_1}(x), \xi_{2,\lambda_2}(x)\right)$ are all of the orders $O(h^2)$. and when $\lambda_1 \neq 0$ and $\lambda_2 \neq 0$, then they are all of $O(h^4)$.

2. The following is a list of the distinct entries in the matrix $E\left((\boldsymbol{s}-\tilde{\boldsymbol{s}})(\boldsymbol{s}-\tilde{\boldsymbol{s}})^T|\boldsymbol{D}\right)$. Based on the calculations we did in the previous section , we can derive the orders of the expectations of these elements, thus leading to a better understanding of the difference between $\hat{g}_n(x;\lambda)$ and $\tilde{g}_n(x;\lambda)$.

Denote (i,j)-th element in the matrix $E\left((\boldsymbol{s}-\tilde{\boldsymbol{s}})(\boldsymbol{s}-\tilde{\boldsymbol{s}})^T|\boldsymbol{D}\right)$ as $a_{ij},\ i,j=1,2,\ldots,5$. Then we can show that

$$
\begin{aligned}
a_{11} &= \frac{1}{2n^2h\sqrt{\pi}}\sum_{i=1}^{n}\phi(x,Z_i,\delta_{0h}^2/2)-\frac{1}{n^2}\sum_{i=1}^{n}\phi^2(x,Z_i,\delta_{0h}^2).\\
a_{12} &= \frac{h}{2n^2\sqrt{\pi}(h^2+2\lambda\sigma_u^2)}\sum_{i=1}^{n}(Z_i-x)\phi(x,Z_i,\delta_{0h}^2/2)-\frac{h^2}{n^2\delta_{0h}^2}\sum_{i=1}^{n}(Z_i-x)\phi^2(x,Z_i,\delta_{0h}^2).\\
a_{13} &= \frac{\lambda h\sigma_u^2}{2n^2\sqrt{\pi}(h^2+2\lambda\sigma_u^2)}\sum_{i=1}^{n}\phi(x,Z_i,\delta_{0h}^2/2)-\frac{h^4}{n^2\delta_{0h}^4}\sum_{i=1}^{n}(x-Z_i)^2\phi^2(x,Z_i,\delta_{0h}^2)\\
&\quad+\frac{h^3}{2n^2\sqrt{\pi}(h^2+2\lambda\sigma_u^2)^2}\sum_{i=1}^{n}(x-Z_i)^2\phi(x,Z_i,\delta_{0h}^2/2)-\frac{\lambda\sigma_u^2h^2}{n^2\delta_{0h}^2}\sum_{i=1}^{n}\phi^2(x,Z_i,\delta_{0h}^2).\\
a_{14} &= \frac{1}{2n^2h\sqrt{\pi}}\sum_{i=1}^{n}Y_i\phi(x,Z_i,\delta_{0h}^2/2)-\frac{1}{n^2}\sum_{i=1}^{n}Y_i\phi^2(x,Z_i,\delta_{0h}^2).\\
a_{15} &= \frac{h}{2n^2\sqrt{\pi}(h^2+2\lambda\sigma_u^2)}\sum_{i=1}^{n}Y_i(Z_i-x)\phi(x,Z_i,\delta_{0h}^2/2)-\frac{h^2}{n^2\delta_{0h}^2}\sum_{i=1}^{n}Y_i(Z_i-x)\phi^2(x,Z_i,\delta_{0h}^2).\\
a_{22} &= \frac{\lambda h\sigma_u^2}{2n^2\sqrt{\pi}(h^2+2\lambda\sigma_u^2)}\sum_{i=1}^{n}\phi(x,Z_i,\delta_{0h}^2/2)-\frac{h^4}{n^2\delta_{0h}^4}\sum_{i=1}^{n}(Z_i-x)^2\phi^2(x,Z_i,\delta_{0h}^2)\\
&\quad+\frac{h^3}{2n^2\sqrt{\pi}(h^2+2\lambda\sigma_u^2)^2}\sum_{i=1}^{n}(x-Z_i)^2\phi(x,Z_i,\delta_{0h}^2/2).\\
a_{23} &= \frac{3\lambda\sigma_u^2h^3}{2n^2\sqrt{\pi}\delta_{0h}^4}\sum_{i=1}^{n}(Z_i-x)\phi(x,Z_i,\delta_{0h}^2/2)+\frac{h^5}{2n^2\sqrt{\pi}\delta_{0h}^6}\sum_{i=1}^{n}(Z_i-x)^3\phi(x,Z_i,\delta_{0h}^2/2)\\
&\quad-\frac{h^6}{n^2\delta_{0h}^6}\sum_{i=1}^{n}(Z_i-x)^3\phi^2(x,Z_i,\delta_{0h}^2)-\frac{\lambda\sigma_u^2h^4}{n^2\delta_{0h}^4}\sum_{i=1}^{n}(Z_i-x)\phi^2(x,Z_i,\delta_{0h}^2).\\
a_{24} &= \frac{h}{2n^2\sqrt{\pi}(h^2+2\lambda\sigma_u^2)}\sum_{i=1}^{n}Y_i(Z_i-x)\phi(x,Z_i,\delta_{0h}^2/2)-\frac{h^2}{n^2\delta_{0h}^2}\sum_{i=1}^{n}Y_i(Z_i-x)\phi^2(x,Z_i,\delta_{0h}^2).\\
a_{25} &= \frac{\lambda h\sigma_u^2}{2n^2\sqrt{\pi}(h^2+2\lambda\sigma_u^2)}\sum_{i=1}^{n}Y_i\phi(x,Z_i,\delta_{0h}^2/2)-\frac{h^4}{n^2\delta_{0h}^4}\sum_{i=1}^{n}Y_i(Z_i-x)^2\phi^2(x,Z_i,\delta_{0h}^2)\\
&\quad+\frac{h^3}{2n^2\sqrt{\pi}(h^2+2\lambda\sigma_u^2)^2}\sum_{i=1}^{n}Y_i(x-Z_i)^2\phi(x,Z_i,\delta_{0h}^2/2).
\end{aligned}
$$

$$a_{33} = \frac{3\lambda^2\sigma_u^4 h^3}{2n^2\sqrt{\pi}(h^2+2\lambda\sigma_u^2)^2}\sum_{i=1}^n \phi(x, Z_i, \delta_{0h}^2/2) - \frac{\lambda^2\sigma_u^4 h^4}{n^2\delta_{0h}^4}\sum_{i=1}^n \phi^2(x, Z_i, \delta_{0h}^2)$$
$$+\frac{3\lambda\sigma_u^2 h^5}{n^2\sqrt{\pi}(h^2+2\lambda\sigma_u^2)^3}\sum_{i=1}^n (Z_i-x)^2\phi(x, Z_i, \delta_{0h}^2/2) - \frac{h^8}{n^2\delta_{0h}^8}\sum_{i=1}^n (Z_i-x)^4\phi^2(x, Z_i, \delta_{0h}^2)$$
$$+\frac{h^7}{2n^2\sqrt{\pi}(h^2+2\lambda\sigma_u^2)^4}\sum_{i=1}^n (Z_i-x)^4\phi(x, Z_i, \delta_{0h}^2/2) - \frac{2\lambda\sigma_u^2 h^6}{n^2\delta_{0h}^6}\sum_{i=1}^n (Z_i-x)^2\phi^2(x, Z_i, \delta_{0h}^2).$$
$$a_{34} = \frac{\lambda h\sigma_u^2}{2n^2\sqrt{\pi}(h^2+2\lambda\sigma_u^2)}\sum_{i=1}^n Y_i\phi(x, Z_i, \delta_{0h}^2/2) - \frac{h^4}{n^2\delta_{0h}^4}\sum_{i=1}^n Y_i(x-Z_i)^2\phi^2(x, Z_i, \delta_{0h}^2)$$
$$+\frac{h^3}{2n^2\sqrt{\pi}(h^2+2\lambda\sigma_u^2)^2}\sum_{i=1}^n Y_i(x-Z_i)^2\phi(x, Z_i, \delta_{0h}^2/2) - \frac{\lambda\sigma_u^2 h^2}{n^2\delta_{0h}^2}\sum_{i=1}^n Y_i\phi^2(x, Z_i, \delta_{0h}^2).$$
$$a_{35} = \frac{3\lambda\sigma_u^2 h^3}{2n^2\sqrt{\pi}\delta_{0h}^4}\sum_{i=1}^n Y_i(Z_i-x)\phi(x, Z_i, \delta_{0h}^2/2) - \frac{h^6}{n^2\delta_{0h}^6}\sum_{i=1}^n Y_i(Z_i-x)^3\phi^2(x, Z_i, \delta_{0h}^2)$$
$$+\frac{h^5}{2n^2\sqrt{\pi}\delta_{0h}^6}\sum_{i=1}^n Y_i(Z_i-x)^3\phi(x, Z_i, \delta_{0h}^2/2) - \frac{\lambda\sigma_u^2 h^4}{n^2\delta_{0h}^4}\sum_{i=1}^n Y_i(Z_i-x)\phi^2(x, Z_i, \delta_{0h}^2).$$
$$a_{44} = \frac{1}{2n^2h\sqrt{\pi}}\sum_{i=1}^n Y_i^2\phi(x, Z_i, \delta_{0h}^2/2) - \frac{1}{n^2}\sum_{i=1}^n Y_i^2\phi^2(x, Z_i, \delta_{0h}^2).$$
$$a_{45} = \frac{h}{2n^2\sqrt{\pi}(h^2+2\lambda\sigma_u^2)}\sum_{i=1}^n Y_i^2(Z_i-x)\phi(x, Z_i, \delta_{0h}^2/2) - \frac{h^2}{n^2\delta_{0h}^2}\sum_{i=1}^n Y_i^2(Z_i-x)\phi^2(x, Z_i, \delta_{0h}^2).$$
$$a_{55} = \frac{\lambda h\sigma_u^2}{2n^2\sqrt{\pi}(h^2+2\lambda\sigma_u^2)}\sum_{i=1}^n Y_i^2\phi(x, Z_i, \delta_{0h}^2/2) - \frac{h^4}{n^2\delta_{0h}^4}\sum_{i=1}^n Y_i^2(Z_i-x)^2\phi^2(x, Z_i, \delta_{0h}^2)$$
$$+\frac{h^3}{2n^2\sqrt{\pi}(h^2+2\lambda\sigma_u^2)^2}\sum_{i=1}^n Y_i^2(x-Z_i)^2\phi(x, Z_i, \delta_{0h}^2/2).$$

3. To show the asymptotic normality of the proposed estimator, we have to check the Lyapunov CLT condition. The following contains the derivations for the asymptotic order of $E|v_{1\lambda}(x)|^3$.

Applying Lemma 7 with $a = 1/3$, $c = \frac{(\lambda+3)\sigma_u^2}{3}$, $m(t) = f_X(t)$, and Lemma 8, we have

$$E\phi^3(x; Z, \delta_{0h}^2) = \iint \phi^3(x; t+u, \delta_{0h}^2)\phi(u; 0, \sigma_u^2)f_X(t)dudt$$
$$= \frac{1}{\delta_{0h}^2}\cdot\frac{1}{2\pi\sqrt{3}}\iint \phi\left(x; t+u, \frac{\delta_{0h}^2}{3}\right))\phi(u; 0, \sigma_u^2)f_X(t)dudt$$
$$= \frac{1}{\delta_{0h}^2}\cdot\frac{1}{2\pi\sqrt{3}}\iint \phi\left(u; x-t, \frac{\delta_{0h}^2}{3}\right))\phi(u; 0, \sigma_u^2)f_X(t)dudt$$

$$= \frac{1}{2\pi\sqrt{3}\delta_{0h}^2} \int \phi\left(t; x, \frac{\delta_{0h}^2}{3} + \sigma_u^2\right) f_X(t)dt = \frac{1}{2\pi\sqrt{3}\delta_{0h}^2} \int \phi\left(t; x, \frac{\delta_{3h}^2}{3}\right) f_X(t)dt$$
$$= \frac{1}{2\pi\sqrt{3}\delta_{0h}^2} \left[\int \phi\left(t; x, \frac{(\lambda+3)\sigma_u^2}{3}\right) f_X(t)dt + \frac{h^2}{6} \int f_X''(t)\phi\left(t-x; 0, \frac{(\lambda+3)\sigma_u^2}{3}\right) dt + o(h^2)\right]$$
$$= \frac{1}{2\pi\sqrt{3}\delta_{0h}^2} \left[f_{0,\frac{\lambda}{3}}(x) + \frac{h^2}{6} f''_{0,\frac{\lambda}{3}}(x) + o(h^2)\right].$$

For $E|Z-x|^3\phi^3(x; Z, \delta_{0h}^2)$, note that

$$E|Z-x|^3\phi^3(x; Z, \delta_{0h}^2) = \iint |t+u-x|^3\phi^3(x; t+u, \delta_{0h}^2)\phi(u; 0, \sigma_u^2)f_X(t)dudt$$
$$= \frac{1}{\delta_{0h}^2} \cdot \frac{1}{2\pi\sqrt{3}} \iint |t+u-x|^3\phi\left(x; t+u, \frac{\delta_{0h}^2}{3}\right))\phi(u; 0, \sigma_u^2)f_X(t)dudt$$
$$= \frac{1}{\delta_{0h}^2} \cdot \frac{1}{2\pi\sqrt{3}} \iint |u-(x-t)|^3\phi\left(u; x-t, \frac{\delta_{0h}^2}{3}\right))\phi(u; 0, \sigma_u^2)f_X(t)dudt$$
$$\leq \frac{4}{2\pi\sqrt{3}\delta_{0h}^2} \iint \left(|u|^3 + |x-t|^3\right)\phi\left(u; x-t, \frac{\delta_{0h}^2}{3}\right)\phi(u; 0, \sigma_u^2)f_X(t)dudt$$
$$= \frac{2}{\pi\sqrt{3}\delta_{0h}^2} \iint |u|^3\phi\left(t; x, \frac{\delta_{3h}^2}{3}\right)\phi\left(u; \frac{\sigma_u^2(x-t)}{\sigma_u^2 + \frac{\delta_{0h}^2}{3}}, \frac{\frac{\sigma_u^2\delta_{0h}^2}{3}}{\sigma_u^2 + \frac{\delta_{0h}^2}{3}}\right) f_X(t)dudt$$
$$+ \frac{2}{\pi\sqrt{3}\delta_{0h}^2} \int |x-t|^3\phi\left(t; x, \frac{\delta_{3h}^2}{3}\right) f_X(t)dt$$
$$= \frac{2}{\pi\sqrt{3}\delta_{0h}^2} \iint |u|^3\phi\left(t; x, \frac{\delta_{3h}^2}{3}\right)\phi\left(u; \frac{3\sigma_u^2(x-t)}{\delta_{3h}^2}, \frac{\sigma_u^2\delta_{0h}^2}{\delta_{3h}^2}\right) f_X(t)dudt$$
$$+ \frac{2}{\pi\sqrt{3}\delta_{0h}^2} \int |x-t|^3\phi\left(t; x, \frac{\delta_{3h}^2}{3}\right) f_X(t)dt$$
$$\leq \frac{2}{\pi\sqrt{3}\delta_{0h}^2} \int \phi\left(t; x, \frac{\delta_{3h}^2}{3}\right)\left[\frac{8\sqrt{2}}{\sqrt{\pi}}\left(\frac{\sigma_u^2\delta_{0h}^2}{\delta_{3h}^2}\right)^{\frac{3}{2}} + 4\left(\frac{3\sigma_u^2|x-t|}{\delta_{3h}^2}\right)^3\right] f_X(t)dt$$
$$+ \frac{2}{\pi\sqrt{3}\delta_{0h}^2} \int |x-t|^3\phi\left(t; x, \frac{\delta_{3h}^2}{3}\right) f_X(t)dt$$
$$= \frac{2}{\pi\sqrt{3}\delta_{0h}^2} \cdot \frac{8\sqrt{2}}{\sqrt{\pi}}\left(\frac{\sigma_u^2\delta_{0h}^2}{\delta_{3h}^2}\right)^{\frac{3}{2}} \int \phi\left(t; x, \frac{\delta_{3h}^2}{3}\right) f_X(t)dt$$
$$+ \frac{2}{\pi\sqrt{3}\delta_{0h}^2} \cdot 4\left(\frac{3\sigma_u^2}{\delta_{3h}^2}\right)^3 \int |x-t|^3\phi\left(t; x, \frac{\delta_{3h}^2}{3}\right) f_X(t)dt$$
$$+ \frac{2}{\pi\sqrt{3}\delta_{0h}^2} \int |x-t|^3\phi\left(t; x, \frac{\delta_{3h}^2}{3}\right) f_X(t)dt.$$

Here, we used the fact that

$$\int |u|^3 \phi(u,\mu,\sigma^2)du = \int |u-\mu+\mu|^3\phi(u,\mu,\sigma^2)du \leq 4\int |u-\mu|^3\phi(u,\mu,\sigma^2)du + 4|\mu|^3$$
$$= 4\sigma^3 \cdot 2^{\frac{3}{2}}\frac{\Gamma(\frac{3+1}{2})}{\sqrt{\pi}} + 4|\mu|^3 = \frac{8\sqrt{2}\sigma^3}{\sqrt{\pi}}\Gamma(2) + 4|\mu|^3 = \frac{8\sqrt{2}\sigma^3}{\sqrt{\pi}} + 4|\mu|^3.$$

For $E|Z-x|^6\phi^3(x;Z,\delta_{0h}^2)$, note that

$$E|Z-x|^6\phi^3(x;Z,\delta_{0h}^2) = \iint |t+u-x|^6\phi^3(x;t+u,\delta_{0h}^2)\phi(u;0,\sigma_u^2)f_X(t)dudt$$
$$= \frac{1}{\delta_{0h}^2}\cdot\frac{1}{2\pi\sqrt{3}}\iint |t+u-x|^6\phi\left(x;t+u,\frac{\delta_{0h}^2}{3}\right))\phi(u;0,\sigma_u^2)f_X(t)dudt$$
$$= \frac{1}{\delta_{0h}^2}\cdot\frac{1}{2\pi\sqrt{3}}\iint |u-(x-t)|^6\phi\left(u;x-t,\frac{\delta_{0h}^2}{3}\right))\phi(u;0,\sigma_u^2)f_X(t)dudt$$
$$\leq \frac{32}{2\pi\sqrt{3}\delta_{0h}^2}\iint \left(|u|^6+|x-t|^6\right)\phi\left(u;x-t,\frac{\delta_{0h}^2}{3}\right)\phi(u;0,\sigma_u^2)f_X(t)dudt$$
$$= \frac{16}{\pi\sqrt{3}\delta_{0h}^2}\iint |u|^6\phi\left(t;x,\frac{\delta_{3h}^2}{3}\right)\phi\left(u;\frac{\sigma_u^2(x-t)}{\sigma_u^2+\frac{\delta_{0h}^2}{3}},\frac{\frac{\sigma_u^2\delta_{0h}^2}{3}}{\sigma_u^2+\frac{\delta_{0h}^2}{3}}\right)f_X(t)dudt$$
$$+\frac{16}{\pi\sqrt{3}\delta_{0h}^2}\int |x-t|^6\phi\left(t;x,\frac{\delta_{3h}^2}{3}\right)f_X(t)dt$$
$$= \frac{16}{\pi\sqrt{3}\delta_{0h}^2}\iint |u|^6\phi\left(t;x,\frac{\delta_{3h}^2}{3}\right)\phi\left(u;\frac{3\sigma_u^2(x-t)}{\delta_{3h}^2},\frac{\sigma_u^2\delta_{0h}^2}{\delta_{3h}^2}\right)f_X(t)dudt$$
$$+\frac{16}{\pi\sqrt{3}\delta_{0h}^2}\int |x-t|^6\phi\left(t;x,\frac{\delta_{3h}^2}{3}\right)f_X(t)dt$$
$$\leq \frac{16}{\pi\sqrt{3}\delta_{0h}^2}\cdot C\cdot\left(\frac{\sigma_u^2\delta_{0h}^2}{\delta_{3h}^2}\right)^3\int \phi\left(t;x,\frac{\delta_{3h}^2}{3}\right)f_X(t)dt$$
$$+\frac{16}{\pi\sqrt{3}\delta_{0h}^2}\cdot 32\cdot\left(\frac{3\sigma_u^2}{\delta_{3h}^2}\right)^6\int |x-t|^6\phi\left(t;x,\frac{\delta_{3h}^2}{3}\right)f_X(t)dt$$
$$+\frac{16}{\pi\sqrt{3}\delta_{0h}^2}\int |x-t|^6\phi\left(t;x,\frac{\delta_{3h}^2}{3}\right)f_X(t)dt,$$

where $C = 480\sigma_u^6$, since $\int |u|^6\phi(u,\mu,\sigma^2)du = \int |u-\mu+\mu|^6\phi(u,\mu,\sigma^2)du$ is bounded above by $32\int |u-\mu|^6\phi(u,\mu,\sigma^2)du + 32\mu^6 = C\sigma^6 + 32\mu^6$. Define $\mu(X) = E\left(|\epsilon|^3\middle|X\right)$. Then,

$$E|Y|^3\phi^3(x;Z,\delta_{0h}^2) = E|g(X)+\epsilon|^3\phi^3(x;Z,\delta_{0h}^2) \leq 4\cdot E|g(X)|^3\phi^3(x;Z,\delta_{0h}^2) + 4\cdot E|\epsilon|^3\phi^3(x;Z,\delta_{0h}^2)$$

$$= 4 \cdot E|g(X)|^3\phi^3(x; Z, \delta_{0h}^2) + 4 \cdot E\delta(X)\phi^3(x; Z, \delta_{0h}^2).$$

Now,

$$E\left[|g(X)|^3\phi^3(x; Z, \delta_{0h}^2)\right] = \iint |g(t)|^3\phi^3(x; t+u, \delta_{0h}^2)\phi(u; 0, \sigma_u^2)f_X(t)dudt$$
$$= \frac{1}{\delta_{0h}^2} \cdot \frac{1}{2\pi\sqrt{3}} \iint |g(t)|^3\phi\left(x; t+u, \frac{\delta_{0h}^2}{3}\right))\phi(u; 0, \sigma_u^2)f_X(t)dudt$$
$$= \frac{1}{\delta_{0h}^2} \cdot \frac{1}{2\pi\sqrt{3}} \iint |g(t)|^3\phi\left(u; x-t, \frac{\delta_{0h}^2}{3}\right))\phi(u; 0, \sigma_u^2)f_X(t)dudt$$
$$= \frac{1}{2\pi\sqrt{3}\delta_{0h}^2} \int |g(t)|^3\phi\left(t; x, \frac{\delta_{3h}^2}{3}\right) f_X(t)dt,$$

therefore,

$$E|Y|^3\phi^3(x; Z, \delta_{0h}^2) \leq \frac{4}{2\pi\sqrt{3}\delta_{0h}^2} \int \left[|g(t)|^3 + \mu(t)\right] \phi\left(t; x, \frac{\delta_{3h}^2}{3}\right) f_X(t)dt.$$

For $E|Y|^3|Z-x|^3\phi^3(x; Z, \delta_{0h}^2)$, note that

$$E|Y|^3|Z-x|^3\phi^3(x; Z, \delta_{0h}^2) \leq 4E\left[|g(X)|^3 + \mu(X)\right] \cdot \left[|Z-x|^3\phi^3(x; Z, \delta_{0h}^2)\right]$$
$$= 4\iint \left[|g(t)|^3 + \mu(t)\right] |t+u-x|^3\phi^3(x; t+u, \delta_{0h}^2)\phi(u; 0, \sigma_u^2)f_X(t)dudt$$
$$\leq 16 \cdot \frac{1}{2\pi\sqrt{3}\delta_{0h}^2} \iint \left[|g(t)|^3 + \mu(t)\right] \left(|u|^3 + |x-t|^3\right) \phi\left(u; x-t, \frac{\delta_{0h}^2}{3}\right) \phi(u; 0, \sigma_u^2)f_X(t)dudt$$
$$= \frac{16}{2\pi\sqrt{3}\delta_{0h}^2} \iint \left[|g(t)|^3 + \mu(t)\right] |u|^3\phi\left(t; x, \frac{\delta_{3h}^2}{3}\right) \cdot$$
$$\phi\left(u; \frac{\sigma_u^2(x-t)}{\sigma_u^2 + \frac{\delta_{0h}^2}{3}}, \frac{\frac{\sigma_u^2\delta_{0h}^2}{3}}{\sigma_u^2 + \frac{\delta_{0h}^2}{3}}\right) f_X(t)dudt$$
$$+ \frac{16}{2\pi\sqrt{3}\delta_{0h}^2} \int \left[|g(t)|^3 + \mu(t)\right] |x-t|^3\phi\left(t; x, \frac{\delta_{3h}^2}{3}\right) f_X(t)dt$$
$$= \frac{16}{2\pi\sqrt{3}\delta_{0h}^2} \iint \left[|g(t)|^3 + \mu(t)\right] |u|^3\phi\left(t; x, \frac{\delta_{3h}^2}{3}\right) \cdot$$
$$\phi\left(u; \frac{3\sigma_u^2(x-t)}{\delta_{3h}^2}, \frac{\sigma_u^2\delta_{0h}^2}{\delta_{3h}^2}\right) f_X(t)dudt$$
$$+ \frac{16}{2\pi\sqrt{3}\delta_{0h}^2} \int \left[|g(t)|^3 + \mu(t)\right] |x-t|^3\phi\left(t; x, \frac{\delta_{3h}^2}{3}\right) f_X(t)dt$$

$$\leq \frac{16}{2\pi\sqrt{3}\delta_{0h}^2} \int \left[|g(t)|^3 + \mu(t)\right] \phi\left(t; x, \frac{\delta_{3h}^2}{3}\right) \cdot$$
$$\left[\frac{8\sqrt{2}}{\sqrt{\pi}} \left(\frac{\sigma_u^2 \delta_{0h}^2}{\delta_{3h}^2}\right)^{\frac{3}{2}} + 4\left(\frac{3\sigma_u^2 |x-t|}{\delta_{3h}^2}\right)^3\right] f_X(t)dt$$
$$+\frac{16}{2\pi\sqrt{3}\delta_{0h}^2} \int \left[|g(t)|^3 + \mu(t)\right] |x-t|^3 \phi\left(t; x, \frac{\delta_{3h}^2}{3}\right) f_X(t)dt$$
$$= \frac{16}{2\pi\sqrt{3}\delta_{0h}^2} \cdot \frac{8\sqrt{2}}{\sqrt{\pi}} \left(\frac{\sigma_u^2 \delta_{0h}^2}{\delta_{3h}^2}\right)^{\frac{3}{2}} \int \left[|g(t)|^3 + \mu(t)\right] \phi\left(t; x, \frac{\delta_{3h}^2}{3}\right) f_X(t)dt$$
$$+\frac{16}{2\pi\sqrt{3}\delta_{0h}^2} \cdot 4\left(\frac{3\sigma_u^2}{\delta_{3h}^2}\right)^3 \int \left[|g(t)|^3 + \mu(t)\right] |x-t|^3 \phi\left(t; x, \frac{\delta_{3h}^2}{3}\right) f_X(t)dt$$
$$+\frac{16}{2\pi\sqrt{3}\delta_{0h}^2} \int \left[|g(t)|^3 + \mu(t)\right] |x-t|^3 \phi\left(t; x, \frac{\delta_{3h}^2}{3}\right) f_X(t)dt.$$

Now, we have to calculate $E(Z-x)^3\phi^2(x, Z, \delta_{0h}^2)$. Recall that for $u \sim N(\mu, \sigma_u^2)$, we have $Eu^3 = 3\mu\sigma_u^2 + \mu^3$, $Eu^4 = 3\sigma^4 + 6\mu^2\sigma_u^2 + \mu^4$. Therefore,

$$E(Z-x)^3\phi^2(x, Z, \delta_{0h}^2) = \iint (t+u-x)^3 \phi^2(u; x-t, \delta_{0h}^2)\phi(u; 0, \sigma_u^2) f_X(t)dudt$$
$$= \frac{1}{2\sqrt{\pi\delta_{0h}^2}} \iint (t+u-x)^3 \phi\left(u; x-t, \frac{\delta_{0h}^2}{2}\right) \phi(u; 0, \sigma_u^2) f_X(t)dudt$$
$$= \frac{1}{2\sqrt{\pi\delta_{0h}^2}} \iint (t+u-x)^3 \phi\left(t, x, \frac{\delta_{2h}^2}{2}\right) \phi\left(u; \frac{2\sigma_u^2(x-t)}{\delta_{2h}^2}, \frac{\sigma_u^2\delta_{0h}^2}{\delta_{2h}^2}\right) f_X(t)dudt$$
$$= \frac{1}{2\sqrt{\pi\delta_{0h}^2}} \int (t-x)^3 \phi\left(t, x, \frac{\delta_{2h}^2}{2}\right) f_X(t)dt - \frac{3}{2\sqrt{\pi\delta_{0h}^2}} \frac{2\sigma_u^2}{\delta_{2h}^2} \int (t-x)^3 \phi\left(t, x, \frac{\delta_{2h}^2}{2}\right) f_X(t)dt$$
$$+\frac{3}{2\sqrt{\pi\delta_{0h}^2}} \int (t-x)\phi\left(t, x, \frac{\delta_{2h}^2}{2}\right) \cdot \left[\left(\frac{2\sigma_u^2(x-t)}{\delta_{2h}^2}\right)^2 + \frac{\sigma_u^2\delta_{0h}^2}{\delta_{2h}^2}\right] f_X(t)dt$$
$$+\frac{1}{2\sqrt{\pi\delta_{0h}^2}} \int \left[3\frac{\sigma_u^2\delta_{0h}^2}{\delta_{2h}^2} \frac{2\sigma_u^2(x-t)}{\delta_{2h}^2} + \left[\frac{2\sigma_u^2(x-t)}{\delta_{2h}^2}\right]^3\right] \phi\left(t, x, \frac{\delta_{2h}^2}{2}\right) f_X(t)dt$$
$$= \frac{1}{2\sqrt{\pi\delta_{0h}^2}} \left[1 - 3\left(\frac{2\sigma_u^2}{\delta_{2h}^2}\right) + 3\left(\frac{2\sigma_u^2}{\delta_{2h}^2}\right)^2 - \left(\frac{2\sigma_u^2}{\delta_{2h}^2}\right)^3\right] \int (t-x)^3 f_X(t)\phi\left(t, x, \frac{\delta_{2h}^2}{2}\right) dt$$
$$+\frac{3}{2\sqrt{\pi\delta_{0h}^2}} \left[1 - \frac{2\sigma_u^2}{\delta_{2h}^2}\right] \frac{\sigma_u^2\delta_{0h}^2}{\delta_{2h}^2} \int (t-x) f_X(t)\phi\left(t, x, \frac{\delta_{2h}^2}{2}\right) dt$$
$$= \frac{1}{2\sqrt{\pi\delta_{0h}^2}} \left[1 - \frac{2\sigma_u^2}{\delta_{2h}^2}\right]^3 \int (t-x)^3 f_X(t)\phi\left(t; x, \frac{\delta_{2h}^2}{2}\right) dt + \frac{3\sigma_u^2}{2\sqrt{\pi\delta_{0h}^2}} \frac{\delta_{0h}^4}{\delta_{2h}^4} \int (t-x) f_X(t)\phi\left(t; x, \frac{\delta_{2h}^2}{2}\right) dt$$
$$= \frac{1}{2\sqrt{\pi\delta_{0h}^2}} \frac{\delta_{0h}^6}{\delta_{2h}^6} \int (t-x)^3 f_X(t)\phi\left(t; x, \frac{\delta_{2h}^2}{2}\right) dt + \frac{3\sigma_u^2}{2\sqrt{\pi\delta_{0h}^2}} \frac{\delta_{0h}^4}{\delta_{2h}^4} \int (t-x) f_X(t)\phi\left(t; x, \frac{\delta_{2h}^2}{2}\right) dt$$

$$
\begin{aligned}
&= \frac{\delta_{0h}^5}{2\sqrt{\pi}\delta_{2h}^6} \int \left[t^3 - 3t^2x + 3tx^2 - x^3\right] f_X(t)\phi\left(t; x, \frac{\delta_{2h}^2}{2}\right) dt + \frac{3\sigma_u^2\delta_{0h}^3}{2\sqrt{\pi}\delta_{2h}^4} \int (t-x) f_X(t)\phi\left(t; x, \frac{\delta_{2h}^2}{2}\right) dt \\
&= \frac{\delta_{0h}^5}{2\sqrt{\pi}\delta_{2h}^6} \left[f_{3,\frac{\lambda}{2}}(x) - 3xf_{2,\frac{\lambda}{2}}(x) + 3x^2 f_{1,\frac{\lambda}{2}}(x) - x^3 f_{0,\frac{\lambda}{2}}(x)\right] \\
&\quad + \frac{\delta_{0h}^5}{2\sqrt{\pi}\delta_{2h}^6} \left[\frac{h^2}{4}\left(f''_{3,\frac{\lambda}{2}}(x) - 3xf''_{2,\frac{\lambda}{2}}(x) + 3x^2 f''_{1,\frac{\lambda}{2}}(x) - x^3 f''_{0,\frac{\lambda}{2}}(x)\right) + o(h^2)\right] \\
&\quad + \frac{3\sigma_u^2\delta_{0h}^3}{2\sqrt{\pi}\delta_{2h}^4} \left[f_{1,\frac{\lambda}{2}}(x) - xf_{0,\frac{\lambda}{2}}(x) + \frac{h^2}{4}\left(f''_{1,\frac{\lambda}{2}}(x) - xf''_{0,\frac{\lambda}{2}}(x)\right) + o(h^2)\right].
\end{aligned}
$$

Similarly, for $EY(Z-x)^3\phi^2(x, Z, \delta_{0h}^2)$ we have,

$$
\begin{aligned}
&E[Y(Z-x)^3\phi^2(x; Z, \delta_{0h}^2)] = E\left((Z-x)^3\phi^2(x; Z, \delta_{0h}^2) E\left[(g(X)+\varepsilon)\Big| X, U\right]\right) \\
&= E\left[g(X)(Z-x)^3\phi^2(x; Z, \delta_{0h}^2)\right] \\
&= \iint g(t)(t+u-x)^3\phi^2(u; x-t, \delta_{0h}^2)\phi(u; 0, \sigma_u^2) f_X(t)\,du\,dt \\
&= \frac{1}{2\sqrt{\pi\delta_{0h}^2}} \iint g(t)(t+u-x)^3\phi\left(u; x-t, \frac{\delta_{0h}^2}{2}\right)\phi(u; 0, \sigma_u^2) f_X(t)\,du\,dt \\
&= \frac{1}{2\sqrt{\pi\delta_{0h}^2}} \iint g(t)(t+u-x)^3\phi\left(t, x, \frac{\delta_{2h}^2}{2}\right)\phi\left(u; \frac{2\sigma_u^2(x-t)}{\delta_{2h}^2}, \frac{\sigma_u^2\delta_{0h}^2}{\delta_{2h}^2}\right) f_X(t)\,du\,dt \\
&= \left[\frac{1}{2\sqrt{\pi\delta_{0h}^2}} - \frac{6\sigma_u^2}{2\sqrt{\pi\delta_{0h}^2}\delta_{2h}^2}\right] \int g(t)(t-x)^3\phi\left(t, x, \frac{\delta_{2h}^2}{2}\right) f_X(t)\,dt \\
&\quad + \frac{3}{2\sqrt{\pi\delta_{0h}^2}} \int g(t)(t-x)\phi\left(t, x, \frac{\delta_{2h}^2}{2}\right)\left[\left(\frac{2\sigma_u^2(x-t)}{\delta_{2h}^2}\right)^2 + \frac{\sigma_u^2\delta_{0h}^2}{\delta_{2h}^2}\right] f_X(t)\,dt \\
&\quad + \frac{1}{2\sqrt{\pi\delta_{0h}^2}} \int g(t)\phi\left(t, x, \frac{\delta_{2h}^2}{2}\right)\left[3\frac{\sigma_u^2\delta_{0h}^2}{\delta_{2h}^2}\frac{2\sigma_u^2(x-t)}{\delta_{2h}^2} + \left[\frac{2\sigma_u^2(x-t)}{\delta_{2h}^2}\right]^3\right] f_X(t)\,dt \\
&= \frac{1}{2\sqrt{\pi\delta_{0h}^2}} \left[1 - 3\left(\frac{2\sigma_u^2}{\delta_{2h}^2}\right) + 3\left(\frac{2\sigma_u^2}{\delta_{2h}^2}\right)^2 - \left(\frac{2\sigma_u^2}{\delta_{2h}^2}\right)^3\right] \int g(t)(t-x)^3 f_X(t)\phi\left(t; x, \frac{\delta_{2h}^2}{2}\right) dt \\
&\quad + \frac{3}{2\sqrt{\pi\delta_{0h}^2}} \left[1 - \frac{2\sigma_u^2}{\delta_{2h}^2}\right] \frac{\sigma_u^2\delta_{0h}^2}{\delta_{2h}^2} \int g(t)(t-x) f_X(t)\phi\left(t; x, \frac{\delta_{2h}^2}{2}\right) dt \\
&= \frac{1}{2\sqrt{\pi\delta_{0h}^2}} \left[1 - \frac{2\sigma_u^2}{\delta_{2h}^2}\right]^3 \int g(t)(t-x)^3 f_X(t)\phi\left(t; x, \frac{\delta_{2h}^2}{2}\right) dt \\
&\quad + \frac{3\sigma_u^2}{2\sqrt{\pi\delta_{0h}^2}} \frac{\delta_{0h}^4}{\delta_{2h}^4} \int g(t)(t-x) f_X(t)\phi\left(t; x, \frac{\delta_{2h}^2}{2}\right) dt \\
&= \frac{\delta_{0h}^5}{2\sqrt{\pi}\delta_{2h}^6} \left[g_{3,\frac{\lambda}{2}}(x) - 3xg_{2,\frac{\lambda}{2}}(x) + 3x^2 g_{1,\frac{\lambda}{2}}(x) - x^3 g_{0,\frac{\lambda}{2}}(x)\right]
\end{aligned}
$$

$$+\frac{\delta_{0h}^5}{2\sqrt{\pi}\delta_{2h}^6}\left[\frac{h^2}{4}\left(g''_{3,\frac{\lambda}{2}}(x)-3xg''_{2,\frac{\lambda}{2}}(x)+3x^2g''_{1,\frac{\lambda}{2}}(x)-x^3g''_{0,\frac{\lambda}{2}}(x)\right)+o(h^2)\right]$$
$$+\frac{3\sigma_u^2\delta_{0h}^3}{2\sqrt{\pi}\delta_{2h}^4}\left[g_{1,\frac{\lambda}{2}}(x)-xg_{0,\frac{\lambda}{2}}(x)+\frac{h^2}{4}\left(g''_{1,\frac{\lambda}{2}}(x)-xg''_{0,\frac{\lambda}{2}}(x)\right)+o(h^2)\right].$$

4. It is well known that the performance of the estimation procedures for bias reduction in the measurement error modelling heavily depends on the signal to noise ratio, or the ratio of σ_x^2 and σ_u^2. Here, we present some simulation results with signal to noise ratios being changed to 40 and 16, that is, we keep $\sigma_u^2 = 0.1$ and 0.25, but change $X \sim N(0,1)$ to $X \sim N(0,2)$. One can see that the performance of all three methods is greatly improved, and sometimes, the naive method provides better results than the SIMEX and EX methods, which is not beyond our expectation, in that such high signal to noise ratio implies the effect of measurement error is nearly negligible.

Table B.1: $g(x) = x\sin(x)$, $X \sim N(0,2)$

σ_u^2	Method		$n=100$		$n=200$		$n=500$	
			MSE	Time(s)	MSE	Time(s)	MSE	Time(s)
0.1	SIMEX	$B=50$	0.038	75.448	0.024	130.337	0.013	302.553
		$B=100$	0.025	147.632	0.023	258.059	0.005	600.734
	EX		0.052	2.365	0.046	3.335	0.015	6.333
	Naive		0.073	0.188	0.046	0.282	0.023	0.603
0.25	SIMEX	$B=50$	0.050	75.198	0.031	130.825	0.011	303.409
		$B=100$	0.027	148.447	0.030	259.758	0.008	604.552
	EX		0.082	2.278	0.046	3.289	0.015	5.837
	Naive		0.133	0.187	0.086	0.288	0.055	0.601

Table B.2: $g(x) = x^2, \quad X \sim N(0, 2)$

σ_u^2	Method		$n = 100$		$n = 200$		$n = 500$	
			MSE	Time(s)	MSE	Time(s)	MSE	Time(s)
0.1	SIMEX	$B = 50$	0.076	75.079	0.012	130.611	0.023	305.085
		$B = 100$	0.053	154.599	0.020	265.856	0.015	613.201
	EX		0.032	2.360	0.093	3.340	0.023	6.326
	Naive		0.060	0.185	0.023	0.295	0.021	0.601
0.25	SIMEX	$B = 50$	0.129	75.778	0.054	134.026	0.041	310.691
		$B = 100$	0.058	148.162	0.046	257.776	0.019	599.158
	EX		0.063	2.281	0.101	3.158	0.019	5.848
	Naive		0.195	0.184	0.080	0.289	0.114	0.608

Table B.3: $g(x) = \exp(x), \quad X \sim N(0, 2)$

σ_u^2	Method		$n = 100$		$n = 200$		$n = 500$	
			MSE	Time(s)	MSE	Time(s)	MSE	Time(s)
0.1	SIMEX	$B = 50$	0.517	75.002	0.029	129.973	0.130	301.813
		$B = 100$	0.425	147.383	0.122	257.704	0.086	600.860
	EX		0.165	2.363	1.204	3.331	0.220	6.330
	Naive		0.072	0.186	0.358	0.288	0.032	0.606
0.25	SIMEX	$B = 50$	0.738	75.285	0.151	130.792	0.184	303.586
		$B = 100$	0.495	148.524	0.260	259.159	0.045	602.798
	EX		0.116	2.270	0.116	2.281	0.151	5.829
	Naive		0.308	0.193	0.368	0.290	0.038	0.608