Online Updating Method to Correct for Measurement Error in Big Data Streams

JooChul Lee, HaiYing Wang, Elizabeth D. Schifano*

Department of Statistics, University of Connecticut, Storrs, CT 06269, USA

Abstract

When huge amounts of data arrive in streams, online updating is an important method to alleviate both computational and data storage issues. The scope of previous research for online updating is extended in the context of the classical linear measurement error model. In the case where some covariates are unknowingly measured with error at the beginning of the stream, but then are measured without error after a particular point along the data stream, the updated estimators ignoring the measurement error are biased for the true parameters. Once the covariates measured without error are first observed, a method to correct the bias of the estimators, as well as to correct the biases in their variance estimator, is proposed; after correction, the traditional online updating method can then proceed as usual. Further, asymptotic distributions for the corrected and updated estimators are established. Simulation studies and a real data analysis with an airline on-time dataset are provided to illustrate the performance of the proposed method.

Keywords: Data compression, errors-in-variables, linear regression, streaming data.

1. Introduction

Continued advances in science and technology have led to a constantly evolving definition of "big data". Regardless of its formal definition, the amounts of data being collected in these fields continue to grow at a remarkably fast pace. Applying statistical models and methods to such big data can cause excessive computational burden, not only in terms of strains on computer memory due to large volume, but also strains in terms of computational efficiency since even seemingly very simple tasks can take an inordinate amount of time to compute [e.g., 19]. To overcome these barriers, statistical and computational methodologies have largely focused on either subsampling-based approaches [e.g., 9, 13, 21, 22], divide-and-conquer approaches [e.g., 12, 6, 16], or online updating approaches [e.g., 15, 19, 20, 28, 29].

The online updating approach for big data analysis is different from the other two approaches since the data is not assumed to exist all at once, but rather arrives sequentially in large chunks from a data stream. In this framework for regression-type analyses, Schifano et al. [15] developed

*Corresponding author: Department of Statistics, University of Connecticut, 215 Glenbrook Road U4120, Storrs, CT 06269, USA; Tel: 1.860.486.3414; Fax: 1.860.486.4113.

Email address: `elizabeth.schifano@uconn.edu` (Elizabeth D. Schifano)

[1] Supplementary materials for this article are available with the electronic version of this manuscript.

[2] R code for implementation is available on GitHub: https://github.com/pedigree07/Measurementerror.

online updating algorithms that update the regression coefficient estimators and their variances as new data arrive; these algorithms are computationally efficient and minimally storage-intensive. Wang et al. [20] expanded the scope of the online updating method by accommodating the arrival of new predictor variables mid-way along the data stream. Furthermore, Wu et al. [28] developed an online updating method for survival analysis under the Cox proportional hazards models, while Xue et al. [29] proposed an online updating-based test to evaluate the proportional hazards assumption.

In this paper, we focus on online updating in the context of the linear errors-in-variables model. Errors-in-variables cause bias in the estimators for the true parameters in statistical models and a loss of power for statistical inference [e.g., 5]. To solve these problems, measurement error models have been discussed extensively under different assumptions and settings: linear models [e.g., 1, 8, 31], generalized linear models [e.g., 17, 2, 10], nonlinear models [e.g., 18, 4, 3, 26, 27], varying-coefficient partially linear models [e.g., 24, 23, 25], and additive partial linear models [e.g., 11].

Unlike previous studies, we assume that the online-updating process begins with a subset of covariates unknowingly measured with error, and then after a particular known point along the data stream, they are measured precisely. Such phenomena may appear in many fields of application with improved instruments for data measurement. For example Sapuppo et al. [14] developed improved instruments for real-time measurement of blood flow velocity, which allows a wider range of velocity measurement than the previous instrument. Similarly, Zhang et al. [30] have improved the psychrometer, which is the sensor for relative humidity measurement, with higher accuracy and stability. Under the online updating framework, the online updated estimators will be biased in general if any of the covariates are measured with error. Once the covariates are no longer measured with error, continuing to naively update the previous estimates (ignoring the measurement error), will also lead to biased estimators for parameters. Thus, we propose to correct the bias of the estimators once the covariates are no longer measured with error, and then proceed with the traditional online updating algorithm after correction. We further derive the asymptotic distribution for the corrected estimators.

The rest of the paper is organized as follows. In Section 2, we briefly review the online updating method in data streams assuming no covariate measurement error, and then propose our method to correct the bias due to covariate measurement error under the linear model framework. In Section 3, simulation studies and real data analysis are conducted. A discussion concludes in Section 4, with technical details provided in the Supplementary Materials.

2. Model and Method

In this section, we first briefly review the online updating method for linear models assuming no covariate measurement error. We then propose an online updating method for linear models to correct for covariate measurement error, assuming after a specific point along the data stream that the covariates are no longer measured with error.

2.1. Online Updating Method

For now, assume that there is no measurement error in the covariates. Suppose that n_k independent observations $\{(y_{ki}, x_{ki}) : i = 1, \ldots, n_k\}$ arrive in blocks, $k = 1, 2, \ldots$, where y_{ki} is the response variable and x_{ki} is the p dimensional covariate vector. Let $\mathbf{y}_k = (y_{k1}, \ldots, y_{kn_k})'$ and $\boldsymbol{\mathcal{X}}_k = [x_{k1}, \ldots, x_{kn_k}]'$, and assume the linear regression model

$$\mathbf{y}_k = \boldsymbol{\mathcal{X}}_k \boldsymbol{\theta} + \boldsymbol{\epsilon}_k, \tag{1}$$

where $\boldsymbol{\theta}$ is a p-dimensional vector and $\boldsymbol{\epsilon}_k = [\epsilon_{k1}, \dots, \epsilon_{kn_k}]'$ is the random error with mean $\mathbf{0}$ and covariance matrix $\sigma^2 \mathbf{I}_{n_k}$.

Schifano et al. [15] formulated an online updating algorithm for estimating cumulative quantities for big stream data. At block k, the cumulative coefficient estimator of $\boldsymbol{\theta}$, based on all data observed to that point, is

$$\hat{\boldsymbol{\theta}}_k = (\mathbf{V}_{k-1} + \boldsymbol{\mathcal{X}}_k' \boldsymbol{\mathcal{X}}_k)^{-1} (\mathbf{V}_{k-1} \hat{\boldsymbol{\theta}}_{k-1} + \boldsymbol{\mathcal{X}}_k' \boldsymbol{\mathcal{X}}_k \hat{\boldsymbol{\vartheta}}_{n_k,k}), \tag{2}$$

where $\mathbf{V}_k = \sum_{j=1}^k \boldsymbol{\mathcal{X}}_j' \boldsymbol{\mathcal{X}}_j$ for $k = 1, 2, \dots$, $\hat{\boldsymbol{\theta}}_0 = \mathbf{0}$, $\mathbf{V}_0 = \mathbf{0}_p$ is the $p \times p$ matrix of zeros, and $\hat{\boldsymbol{\vartheta}}_{n_k,k}$ is the least squares estimator obtained from the k^{th} block. At the final accumulation point, the cumulative coefficient estimator is the same as the divide-and-recombine estimator [e.g., 12], and in the special case of a linear model, both of these estimators also coincide with the full-data least squares estimator, if it could be obtained with a super-computer. The sum of squared error (SSE) based on the cumulative data at block k is

$$\text{SSE}_k = \text{SSE}_{k-1} + \text{SSE}_{n_k,k} + \hat{\boldsymbol{\theta}}_{k-1}' \mathbf{V}_{k-1} \hat{\boldsymbol{\theta}}_{k-1} + \hat{\boldsymbol{\vartheta}}_{n_k,k}' \boldsymbol{\mathcal{X}}_k' \boldsymbol{\mathcal{X}}_k \hat{\boldsymbol{\vartheta}}_{n_k,k} - \hat{\boldsymbol{\theta}}_k' \mathbf{V}_k \hat{\boldsymbol{\theta}}_k, \tag{3}$$

where $\text{SSE}_{n_k,k}$ is the SSE from the k^{th} block. The corresponding mean squared errors (MSE) are $\text{MSE}_{n_k,k} = \text{SSE}_{n_k,k}/(n_k - p)$ and $\text{MSE}_k = \text{SSE}_k/(N_k - p)$, where $N_k = \sum_{j=1}^k n_j$.

As can be observed from equations (2) and (3), online updating for estimation and inference requires only quantities ($\hat{\boldsymbol{\vartheta}}_{n_k,k}$, $\text{SSE}_{n_k,k}$, $\boldsymbol{\mathcal{X}}_k' \boldsymbol{\mathcal{X}}_k, n_k$) based on the current data ($\mathbf{y}_k, \boldsymbol{\mathcal{X}}_k$), and cumulative quantities ($\hat{\boldsymbol{\theta}}_{k-1}$, SSE_{k-1}, $\mathbf{V}_{k-1}$, N_{k-1}) computed at the previous accumulation point. Thus, the online-updating method is advantageous in this respect, as it does not require storage of historical data.

2.2. Online Updating at the Change of Measurement

As noted earlier, some covariates may actually be measured with error. Suppose we partition $\boldsymbol{\mathcal{X}}_k$ from equation (1) as $\boldsymbol{\mathcal{X}}_k = (\mathbf{Z}_k, \mathbf{X}_k)$, where $\mathbf{Z}_k = [\mathbf{z}_{k1}, \dots, \mathbf{z}_{kn_k}]'$ are not measured with error, but $\mathbf{X}_k = [\mathbf{x}_{k1}, \dots, \mathbf{x}_{kn_k}]'$ are initially measured with error. We can rewrite equation (1) as

$$\mathbf{y}_k = \mathbf{Z}_k \boldsymbol{\alpha} + \mathbf{X}_k \boldsymbol{\beta} + \boldsymbol{\epsilon}_k, \tag{4}$$

where $\boldsymbol{\theta}$ from equation (1) is partitioned correspondingly as $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\beta}')'$ with $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ being p_1- and p_2-dimensional vectors, respectively. Suppose for blocks k, $k = 1, \dots K$, we observe $\mathbf{W}_k$ instead of $\mathbf{X}_k$, where $\mathbf{W}_k$ follows the classical measurement error model, $\mathbf{W}_k = \mathbf{X}_k + \mathbf{U}_k$ and $\mathbf{U}_k = [\mathbf{u}_{k1}, \dots, \mathbf{u}_{kn_k}]'$ is the matrix associated with measurement error where $\mathbf{u}_{ki}$ has mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_u$ for $1 \le i \le n_k$. We assume that $\mathbf{u}_{ki}, \epsilon_{ki}$ and $(\mathbf{z}_{ki}, \mathbf{x}_{ki})$ are mutually independent for all k. For simplicity, we assume that $(\mathbf{z}_{ki}, \mathbf{x}_{ki}, \mathbf{u}_{ki})$ are independent and identically distributed each with the same joint distribution as $(\mathbf{z}, \mathbf{x}, \mathbf{u})$.

We assume for any $k \le K$, the block-wise estimator $\tilde{\boldsymbol{\vartheta}}_{n_k,k} = (\tilde{\mathbf{a}}_{n_k,k}', \tilde{\mathbf{b}}_{n_k,k}')'$ based on the k^{th} dataset is obtained by minimizing the following criterion

$$\tilde{\boldsymbol{\vartheta}}_{n_k,k} = \arg \min_{(\mathbf{a}', \mathbf{b}')'} \| \mathbf{y}_k - \mathbf{Z}_k \mathbf{a} - \mathbf{W}_k \mathbf{b} \|^2,$$

and cumulative coefficient estimator $\tilde{\boldsymbol{\theta}}_k = (\tilde{\mathbf{a}}_k', \tilde{\mathbf{b}}_k')'$, and SSE based on the cumulative data are

$$\tilde{\boldsymbol{\theta}}_k = (\widetilde{\mathbf{V}}_{k-1} + \widetilde{\boldsymbol{\mathcal{X}}}_k' \widetilde{\boldsymbol{\mathcal{X}}}_k)^{-1} (\widetilde{\mathbf{V}}_{k-1} \tilde{\boldsymbol{\theta}}_{k-1} + \widetilde{\boldsymbol{\mathcal{X}}}_k' \widetilde{\boldsymbol{\mathcal{X}}}_k \tilde{\boldsymbol{\vartheta}}_{n_k,k}) \text{ and}$$

$$\widetilde{\text{SSE}}_k = \widetilde{\text{SSE}}_{k-1} + \widetilde{\text{SSE}}_{n_K,K} + \tilde{\boldsymbol{\theta}}'_{k-1}\widetilde{\mathbf{V}}_{k-1}\tilde{\boldsymbol{\theta}}_{k-1} + \tilde{\boldsymbol{\vartheta}}'_{n_k,k}\widetilde{\boldsymbol{\mathcal{X}}}'_k\widetilde{\boldsymbol{\mathcal{X}}}_k\tilde{\boldsymbol{\vartheta}}_{n_k,k} - \tilde{\boldsymbol{\theta}}'_k\widetilde{\mathbf{V}}_k\tilde{\boldsymbol{\theta}}_k,$$

respectively, where $\widetilde{\boldsymbol{\mathcal{X}}}_k = (\mathbf{Z}_k, \mathbf{W}_k)$, $\widetilde{\mathbf{V}}_k = \sum_{j=1}^k \widetilde{\boldsymbol{\mathcal{X}}}'_j\widetilde{\boldsymbol{\mathcal{X}}}_j$, $\tilde{\boldsymbol{\theta}}_{k-1} = \mathbf{0}$, and $\widetilde{\mathbf{V}}_0 = \mathbf{0}_p$.

Note that in the online-updating setting for linear models, cumulative estimator $\tilde{\boldsymbol{\theta}}_k$ is exactly the same as the estimator resulting from fitting the all the data from blocks 1 through K simultaneously via least-squares. It is also well known that in the classical measurement error setting, estimators resulting from models with errors-in-covariates are biased, and the estimated variances are also incorrect [e.g., 5].

If, after block K, the covariates in $\mathbf{X}_k$ are no longer measured with error, i.e., $\mathbf{X}_k$ is available instead of $\mathbf{W}_k$ for $k > K$, we propose to correct the bias in the previous cumulative estimators $\tilde{\boldsymbol{\theta}}_k$ and $\widetilde{\text{SSE}}_k$, based on covariates in $\mathbf{W}_k$, before proceeding with the online-updating process. The steps are described in the following subsections. For ease of readability, we summarize here some notation that will be used in this paper. We use $\hat{\boldsymbol{\vartheta}}$ for coefficient estimators from current data block based on true covariates ($\boldsymbol{\mathcal{X}}_k$), $\tilde{\boldsymbol{\vartheta}}$ for coefficient estimator from current data block based on some covariates measured with error ($\widetilde{\boldsymbol{\mathcal{X}}}_k$), $\hat{\boldsymbol{\theta}}$ for cumulative coefficient estimators based on the true covariates ($\boldsymbol{\mathcal{X}}_k$), $\tilde{\boldsymbol{\theta}}$ for cumulative coefficient estimators based on some covariates measured with error ($\widetilde{\boldsymbol{\mathcal{X}}}_k$), $\tilde{\boldsymbol{\theta}}^c$ for cumulative coefficient estimators with bias correction of $\tilde{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\theta}}^c$ for updated cumulative coefficient estimators after bias correction. For sum of squared errors, we use $\widetilde{\text{SSE}}$ for sum of squared errors when some covariates are measured with error ($\widetilde{\boldsymbol{\mathcal{X}}}_k$), $\widetilde{\text{SSE}}^c$ for sum of squared errors with bias correction of $\widetilde{\text{SSE}}$, and $\widehat{\text{SSE}}^c$ for updated sum of squared errors after bias correction.

2.2.1. Online Updating at the Change of Measurement

At block $K+1$ where $\mathbf{X}_{K+1}$ is first observed, we correct the bias occurring from the measurement error before updating the estimators. In other words, we wish to approximate $\hat{\boldsymbol{\theta}}_K = (\hat{\boldsymbol{\alpha}}'_K, \hat{\boldsymbol{\beta}}'_K)'$ and SSE_K, the cumulative coefficient estimator from equation (2) and SSE from equation (3) at the accumulation point $k = K$ *assuming no measurement error*, before updating with the current information measured without error in block $K+1$. As before, write $\mathbf{V}_K = \sum_{j=1}^K \boldsymbol{\mathcal{X}}'_j\boldsymbol{\mathcal{X}}_j$. Since $\sum_{j=1}^K \mathbf{W}'_j\mathbf{y}_j = \sum_{j=1}^K \mathbf{X}'_j\mathbf{y}_j + o_p(N_K)$, $\hat{\boldsymbol{\theta}}_K$ is approximated as

$$\hat{\boldsymbol{\theta}}_K = (\mathbf{V}_K)^{-1}\sum_{j=1}^K \boldsymbol{\mathcal{X}}'_j\mathbf{y}_j = (\mathbf{V}_K)^{-1}\widetilde{\mathbf{V}}_K\tilde{\boldsymbol{\theta}}_K + o_p(1).$$

However, $\mathbf{V}_K$ cannot be calculated directly because $\mathbf{X}_k$ is not observable for $k \leq K$. Note that for large N_K and n_{K+1}, $\sum_{j=1}^K \mathbf{X}'_j\mathbf{X}_j/N_K \approx \mathbf{X}'_{K+1}\mathbf{X}_{K+1}/n_{K+1}$ where n_{K+1} is the number of observations in block $K+1$. Also, $\sum_{j=1}^K \mathbf{Z}'_j\mathbf{W}_j = \sum_{j=1}^K \mathbf{Z}'_j\mathbf{X}_j + o_p(N_K)$. Thus, we replace $\sum_{j=1}^K \mathbf{X}'_j\mathbf{X}_j$ and $\sum_{j=1}^K \mathbf{Z}'_j\mathbf{X}_j$ by $\frac{N_K}{n_{K+1}}\mathbf{X}'_{K+1}\mathbf{X}_{K+1}$ and $\sum_{j=1}^K \mathbf{Z}'_j\mathbf{W}_j$, respectively. Then, the (biased) cumulative coefficient estimator for $\boldsymbol{\theta}$ at accumulation point $k = K$, i.e., $\tilde{\boldsymbol{\theta}}_K$, can be corrected to approximate $\hat{\boldsymbol{\theta}}_K$ as

$$\tilde{\boldsymbol{\theta}}^c_K = (\widetilde{\mathbf{V}}^c_K)^{-1}\widetilde{\mathbf{V}}_K\tilde{\boldsymbol{\theta}}_K, \tag{5}$$

where $\widetilde{\mathbf{V}}^c_K = \widetilde{\mathbf{V}}_K - \mathbf{T}$,

$$\mathbf{T} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & t^* \end{bmatrix}, \text{ and } t^* = \sum_{j=1}^K \mathbf{W}'_j\mathbf{W}_j - (N_K/n_{K+1})\mathbf{X}'_{K+1}\mathbf{X}_{K+1}. \tag{6}$$

We have the following theorem describing asymptotic properties of $\tilde{\boldsymbol{\theta}}_K^c$. The proof is in Appendix A of the Supplementary Material.

Theorem 1. *Assume that* $\mathbf{E}\|\mathbf{z}\|^2 < \infty$ *and* $\mathbf{E}\|\mathbf{x}\|^4 < \infty$. *Suppose* $\mathbf{u}, \epsilon$ *and* $(\mathbf{z}, \mathbf{x})$ *are mutually independent and* $\mathbf{V} = \mathbf{E}\begin{bmatrix}\mathbf{zz}' & \mathbf{zx}' \\ \mathbf{xz}' & \mathbf{xx}'\end{bmatrix}$ *is positive definite. When* $N_K, n_{K+1} \to \infty$, *we have the following results for* $\tilde{\boldsymbol{\theta}}_K^c$ *defined in equation* (5).
Case 1. If $n_{K+1}/N_K \to 0$, *then*

$$\sqrt{n_{K+1}}(\tilde{\boldsymbol{\theta}}_K^c - \boldsymbol{\theta}) \xrightarrow{d} N(0, \mathbf{V}^{-1}\mathbf{F}_1\mathbf{V}^{-1}); \tag{7}$$

Case 2. if $N_K/n_{K+1} \to 0$, *then*

$$\sqrt{N_K}(\tilde{\boldsymbol{\theta}}_K^c - \boldsymbol{\theta}) \xrightarrow{d} N(0, \mathbf{V}^{-1}\mathbf{F}_2\mathbf{V}^{-1}); \tag{8}$$

and Case 3. if $N_K/n_{K+1} \to h$ *for some constant* $0 < h < \infty$, *then*

$$\sqrt{N_K}(\tilde{\boldsymbol{\theta}}_K^c - \boldsymbol{\theta}) \xrightarrow{d} N(0, \mathbf{V}^{-1}\mathbf{F}_3\mathbf{V}^{-1}); \tag{9}$$

here, $\xrightarrow{d}$ *denotes convergence in distribution,* $\mathbf{F}_1 = \begin{bmatrix}\mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Var}(\mathbf{xx}'\boldsymbol{\beta})\end{bmatrix}$, $\mathbf{F}_2 = \begin{bmatrix}f_{11} & f_{12} \\ f_{12}' & f_{22}\end{bmatrix}$, $\mathbf{F}_3 = \mathbf{F}_2 + h\mathbf{F_1}$, $f_{11} = \mathbf{E}(\mathbf{zz}')(\sigma^2\mathbf{I} + \boldsymbol{\beta}'\boldsymbol{\Sigma}_u\boldsymbol{\beta}), f_{12} = \mathbf{E}(\mathbf{zx}')(\sigma^2\mathbf{I} - \boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{\Sigma}_u)$, *and* $f_{22} = \mathbf{Var}(\mathbf{xx}'\boldsymbol{\beta}) + (\mathbf{E}(\mathbf{xx}') + \boldsymbol{\Sigma}_u)\sigma^2 + \boldsymbol{\Sigma}_u\boldsymbol{\beta}'\mathbf{E}(\mathbf{xx}')\boldsymbol{\beta}$.

Remark 1. Since $\mathbf{V}^{-1}\mathbf{F}_1\mathbf{V}^{-1}$ in (7) is not of full rank, there exists a $p_1 \times p$ matrix $\mathbf{M}$ such that $\sqrt{n_{K+1}}\mathbf{M}(\tilde{\boldsymbol{\theta}}_K^c - \boldsymbol{\theta}) = o_p(1)$ in Case 1 of Theorem 1. For this case, $\begin{bmatrix}\mathbf{I}_{p_1} & \mathbf{0}\end{bmatrix}(\tilde{\mathbf{V}}_K^c/N_K)$ is a particular choice of $\mathbf{M}$ considered to derive the asymptotic normality. Since $\begin{bmatrix}\mathbf{I}_{p_1} & \mathbf{0}\end{bmatrix}(\tilde{\mathbf{V}}_K^c/N_K)(\tilde{\boldsymbol{\theta}}_K^c - \boldsymbol{\theta}) = O_p(1/\sqrt{N_K})$, we obtain that $\sqrt{N_K}\begin{bmatrix}\mathbf{I}_{p_1} & \mathbf{0}\end{bmatrix}(\tilde{\mathbf{V}}_K^c/N_K)(\tilde{\boldsymbol{\theta}}_K^c - \boldsymbol{\theta})$ converges in distribution to $N(\mathbf{0}, f_{11})$. However, $\tilde{\mathbf{V}}_K^c/N_K$ cannot be replaced by $\mathbf{V}$ because $\sqrt{N_K}\begin{bmatrix}\mathbf{I}_{p_1} & \mathbf{0}\end{bmatrix}\mathbf{V}(\tilde{\boldsymbol{\theta}}_K^c - \boldsymbol{\theta})$ does not converge in distribution under the condition in Case 1 of Theorem 1. The detailed derivation of the asymptotic normality is in Appendix B.

Now, we discuss SSE and wish to correct the bias from $\widetilde{\mathrm{SSE}}_K$. SSE_K can be expressed as

$$\begin{aligned}\mathrm{SSE}_K &= \sum_{j=1}^{K}\|\mathbf{y}_j - \boldsymbol{\mathcal{X}}_j\hat{\boldsymbol{\theta}}_K\|^2 \\ &= \sum_{j=1}^{K}\|\mathbf{y}_j - \tilde{\boldsymbol{\mathcal{X}}}_j\tilde{\boldsymbol{\theta}}_K\|^2 + \sum_{j=1}^{K}\|\tilde{\boldsymbol{\mathcal{X}}}_j\tilde{\boldsymbol{\theta}}_K - \boldsymbol{\mathcal{X}}_j\hat{\boldsymbol{\theta}}_K\|^2 + 2\sum_{j=1}^{K}(\mathbf{y}_j - \tilde{\boldsymbol{\mathcal{X}}}_j\tilde{\boldsymbol{\theta}}_K)'(\tilde{\boldsymbol{\mathcal{X}}}_j\tilde{\boldsymbol{\theta}}_K - \boldsymbol{\mathcal{X}}_j\hat{\boldsymbol{\theta}}_K) \\ &= \widetilde{\mathrm{SSE}}_K - \sum_{j=1}^{K}\mathbf{A}_j'\mathbf{A}_j + 2\sum_{j=1}^{K}(\mathbf{y}_j - \boldsymbol{\mathcal{X}}_j\hat{\boldsymbol{\theta}}_K)'\mathbf{A}_j,\end{aligned} \tag{10}$$

where $\mathbf{A}_j = \tilde{\boldsymbol{\mathcal{X}}}_j\tilde{\boldsymbol{\theta}}_K - \boldsymbol{\mathcal{X}}_j\hat{\boldsymbol{\theta}}_K$. Since $\sum_{j=1}^K \mathbf{A}_j'\mathbf{A}_j = \sum_{j=1}^K(\tilde{\boldsymbol{\theta}}_K - \hat{\boldsymbol{\theta}}_K)'\boldsymbol{\mathcal{X}}_j'\boldsymbol{\mathcal{X}}_j(\tilde{\boldsymbol{\theta}}_K - \hat{\boldsymbol{\theta}}_K) + \tilde{\mathbf{b}}_K'\sum_{j=1}^K(\mathbf{W}_j'\mathbf{W}_j - \mathbf{X}_j'\mathbf{X}_j)\tilde{\mathbf{b}}_K + o_p(N_K)$, $2\sum_{j=1}^K(\mathbf{y}_j - \boldsymbol{\mathcal{X}}_j\hat{\boldsymbol{\theta}}_K)'\mathbf{A}_j = o_p(N_K)$ in the last two terms in equation (10), and

$\hat{\boldsymbol{\theta}}_K = \tilde{\boldsymbol{\theta}}_K^c + o_P(1)$. SSE_K is approximated as

$$\text{SSE}_K \approx \widetilde{\text{SSE}}_K - \boldsymbol{\delta}_k' \mathbf{V}_K \boldsymbol{\delta}_K - \tilde{\mathbf{b}}_K' \sum_{j=1}^{K} \left(\mathbf{W}_j' \mathbf{W}_j - \mathbf{X}_j' \mathbf{X}_j \right) \tilde{\mathbf{b}}_K, \tag{11}$$

where $\boldsymbol{\delta}_K = \tilde{\boldsymbol{\theta}}_K - \tilde{\boldsymbol{\theta}}_K^c$. The detailed derivation of the approximation to SSE_K is in Appendix C.

Since we cannot directly obtain $\mathbf{V}_K$ and $\sum_{j=1}^{K} \mathbf{X}_j' \mathbf{X}_j$ on the right-hand side in equation (11), they are replaced with $\widetilde{\mathbf{V}}_K^c$ and $(N_K/n_{K+1})\mathbf{X}_{K+1}' \mathbf{X}_{K+1}$, respectively. Thus, we propose a corrected SSE:

$$\widetilde{\text{SSE}}_K^c = \widetilde{\text{SSE}}_K - \boldsymbol{\delta}_K' \widetilde{\mathbf{V}}_K^c \boldsymbol{\delta}_K - \tilde{\boldsymbol{\theta}}_K' \mathbf{T} \tilde{\boldsymbol{\theta}}_K,$$

where $\mathbf{T}$ is defined as in (6).

2.2.2. Updating estimates

The updated estimates at block $K+1$ can be calculated with the quantities after the bias correction, $(\tilde{\boldsymbol{\theta}}_K^c, \widetilde{\text{SSE}}_K^c, \widetilde{\mathbf{V}}_K^c)$. Using these quantities and equation (2), we propose a cumulative coefficient estimator at the accumulation point $K+1$,

$$\hat{\boldsymbol{\theta}}_{K+1}^c = (\widetilde{\mathbf{V}}_K^c + \boldsymbol{\mathcal{X}}_{K+1}' \boldsymbol{\mathcal{X}}_{K+1})^{-1} [\widetilde{\mathbf{V}}_K^c \tilde{\boldsymbol{\theta}}_K^c + \boldsymbol{\mathcal{X}}_{K+1}' \boldsymbol{\mathcal{X}}_{K+1} \hat{\boldsymbol{\vartheta}}_{n_{K+1},K+1}], \tag{12}$$

where $\hat{\boldsymbol{\vartheta}}_{n_{K+1},K+1} = (\hat{\boldsymbol{\alpha}}_{n_{K+1},K+1}', \hat{\boldsymbol{\beta}}_{n_{K+1},K+1}')'$ is the least squares estimator in the model (4) based on $(K+1)^{th}$ dataset. The following result illustrates the asymptotic properties of the updated estimator $\hat{\boldsymbol{\theta}}_{K+1}^c$.

Theorem 2. *Assume that* $\mathbf{E}\|\mathbf{z}\|^2 < \infty$ *and* $\mathbf{E}\|\mathbf{x}\|^4 < \infty$*. Suppose* $\mathbf{u}, \epsilon$ *and* $(\mathbf{z}, \mathbf{x})$ *are mutually independent and* $\mathbf{V} = \mathbf{E}\begin{bmatrix} \mathbf{z}\mathbf{z}' & \mathbf{z}\mathbf{x}' \\ \mathbf{x}\mathbf{z}' & \mathbf{x}\mathbf{x}' \end{bmatrix}$ *is positive definite. When* $N_K, n_{K+1} \to \infty$*, we have the following results of* $\hat{\boldsymbol{\theta}}_{K+1}^c$ *in equation* (12).
Case 1. If $n_{K+1}/N_K \to 0$*, then*

$$\sqrt{n_{K+1}}(\hat{\boldsymbol{\theta}}_{K+1}^c - \boldsymbol{\theta}) \xrightarrow{d} N(0, \mathbf{V}^{-1}\mathbf{F}_1\mathbf{V}^{-1}); \tag{13}$$

Case 2. if $N_K/n_{K+1} \to 0$*, then*

$$\sqrt{N_{K+1}}(\hat{\boldsymbol{\theta}}_{K+1}^c - \boldsymbol{\theta}) \xrightarrow{d} N(0, \mathbf{V}^{-1}\mathbf{F}_4\mathbf{V}^{-1}); \tag{14}$$

and Case 3. if $N_K/n_{K+1} \to h$ *for some constant* $0 < h < \infty$*, then*

$$\sqrt{N_{K+1}}(\hat{\boldsymbol{\theta}}_{K+1}^c - \boldsymbol{\theta}) \xrightarrow{d} N(0, \mathbf{V}^{-1}\mathbf{F}_5\mathbf{V}^{-1}); \tag{15}$$

where $\mathbf{F}_4 = \mathbf{V}\sigma^2$*,* $\mathbf{F}_5 = (h^2\mathbf{F_1} + \mathbf{V}\sigma^2)/(h+1) + [h/(h+1)]\mathbf{F_2}$*, and* $\mathbf{F_1}, \mathbf{F_2}$ *are the same as those in Theorem 1.*

For the asymptotic variance in Theorem 2, we first discuss SSE. Using equation (3) and the quantities with the bias correction, SSE at the accumulation point $K+1$ can be estimated as

$$\begin{aligned} \widehat{\text{SSE}}_{K+1}^c \quad = \quad & \widetilde{\text{SSE}}_K^c + \text{SSE}_{n_{K+1},K+1} + (\tilde{\boldsymbol{\theta}}_K^c)' \widetilde{\mathbf{V}}_K^c \tilde{\boldsymbol{\theta}}_K^c + \hat{\boldsymbol{\vartheta}}_{n_{K+1},K+1}' \boldsymbol{\mathcal{X}}_{K+1}' \boldsymbol{\mathcal{X}}_{K+1} \hat{\boldsymbol{\vartheta}}_{n_{K+1},K+1} \\ & - \hat{\boldsymbol{\theta}}_{K+1}^c (\widetilde{\mathbf{V}}_K^c + \boldsymbol{\mathcal{X}}_{K+1}' \boldsymbol{\mathcal{X}}_{K+1}) \hat{\boldsymbol{\theta}}_{K+1}^c. \end{aligned}$$

Note that when $n_{K+1} \to \infty$, $\boldsymbol{\mathcal{X}}'_{K+1}\boldsymbol{\mathcal{X}}_{K+1}/n_{K+1}$ converges in probability to $\mathbf{V}$. From the proof of Theorem 2 (see Appendix D), an estimator for the asymptotic variance of $\hat{\boldsymbol{\theta}}^c_{K+1}$ is $\hat{\mathbf{V}}^{-1}\hat{\mathbf{F}}\hat{\mathbf{V}}^{-1}$, where $\hat{\mathbf{V}} = (\widetilde{\mathbf{V}}^c_K + \boldsymbol{\mathcal{X}}'_{K+1}\boldsymbol{\mathcal{X}}_{K+1})/N_{K+1}$,

$$\begin{aligned}\hat{\mathbf{F}} &= N_{K+1}^{-1}\Big\{(n_{K+1}/N_{K+1})\Big((N_K/n_{K+1})^2\,\hat{\mathbf{F}}_1 + (N_{K+1}-p)^{-1}\widehat{\text{SSE}}^c_{K+1}\hat{\mathbf{V}}\Big)\\ &\quad + (N_K/N_{K+1})\,\hat{\mathbf{F}}_2\Big\},\end{aligned} \tag{16}$$

where

$$\hat{\mathbf{F}}_1 = n_{K+1}^{-1}\sum_{i=1}^{n_{K+1}}\left(\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \left(\mathbf{x}_{K+1,i}\mathbf{x}'_{K+1,i} - \widehat{\mathbf{E}(\mathbf{x}\mathbf{x}')}\right)\hat{\boldsymbol{\beta}}_{n_{K+1},K+1}\end{bmatrix}\right)^{\otimes 2},$$

$$\hat{\mathbf{F}}_2 = N_K^{-1}\sum_{j=1}^{K}\sum_{i=1}^{n_j}\left(\begin{bmatrix}\mathbf{z}_{ji}\\ \mathbf{w}_{ji}\end{bmatrix}y_{ji} - \begin{bmatrix}\mathbf{z}_{ji}\mathbf{z}'_{ji} & \mathbf{z}_{ji}\mathbf{w}'_{ji}\\ \mathbf{w}_{ji}\mathbf{z}'_{ji} & \widehat{\mathbf{E}(\mathbf{x}\mathbf{x}')}\end{bmatrix}\hat{\boldsymbol{\vartheta}}_{n_{K+1},K+1}\right)^{\otimes 2},$$

$\mathbf{G}^{\otimes 2} = \mathbf{G}\mathbf{G}'$ for any matrix $\mathbf{G}$, and $\widehat{\mathbf{E}(\mathbf{x}\mathbf{x}')} = \sum_{i=1}^{n_{K+1}}\mathbf{x}_{K+1,i}\mathbf{x}'_{K+1,i}/n_{K+1}$.

2.2.3. Continue updating

We now discuss the cumulative coefficient estimator after block $K+1$. By the online updating algorithm, the cumulative coefficient estimator at the accumulation point $s > K+1$ is

$$\hat{\boldsymbol{\theta}}^c_s = (\widehat{\mathbf{V}}^c_{s-1} + \boldsymbol{\mathcal{X}}'_s\boldsymbol{\mathcal{X}}_s)^{-1}(\widehat{\mathbf{V}}^c_{s-1}\hat{\boldsymbol{\theta}}^c_{s-1} + \boldsymbol{\mathcal{X}}'_s\boldsymbol{\mathcal{X}}_s\hat{\boldsymbol{\vartheta}}_{n_s,s}), \tag{17}$$

where $\widehat{\mathbf{V}}^c_s = \widehat{\mathbf{V}}^c_{s-1} + \boldsymbol{\mathcal{X}}'_s\boldsymbol{\mathcal{X}}_s$. In a data stream where new data keep coming in, it is natural that the size of the accumulated data since block $K+1$ become much bigger than the size of the prior data blocks. Thus, we study the estimator for the accumulation point $s > K+1$ such that $N_K/\sum_{j=K+1}^{s} n_j \to 0$. For this scenario, we obtain the following result in the same way as Case 2 in Theorem 2,

$$\sqrt{N_s}(\hat{\boldsymbol{\theta}}^c_s - \boldsymbol{\theta}) \xrightarrow{d} N(0, \mathbf{V}^{-1}\sigma^2). \tag{18}$$

The variance-covariance matrix in (18) can be estimated by $(\widehat{\mathbf{V}}^c_s/N_s)\widehat{\text{MSE}}^c_s$ where $\widehat{\text{MSE}}^c_s = \widehat{\text{SSE}}^c_s/(N_s - p)$.

3. Numerical Study

3.1. Simulation Study

In the simulation, we consider two blocks of data with sizes n_1 and n_2, respectively. The data is generated by a linear regression model with

$$y_i = \alpha_0 + \mathbf{z}'_i\boldsymbol{\alpha} + \mathbf{x}'_i\boldsymbol{\beta} + \varepsilon_i, \quad i = 1,\ldots,n, \tag{19}$$

where ε_i's are uncorrelated error terms and follow a normal distribution with mean zero and variance σ^2, and $n = n_1 + n_2$. Covariates $(\mathbf{z}_i' \ \mathbf{x}_i')'$ are generated from a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ with σ_{zx} as off-diagonal entries and 1 as diagonal entries. At the first block of data, the covariates with measurement error are $\mathbf{w}_i = \mathbf{x}_i + \mathbf{u}_i$ where $\mathbf{u}_i \sim N(\mathbf{0}, \Sigma_u)$ for $i = 1, \ldots, n_1$ and $\Sigma_u = \sigma_u^2 0.1^{I(i \neq j)}$; the covariates in the second block are not measured with error. To generate the covariates, we consider $\boldsymbol{\mu} \in \{\mathbf{0}, \mathbf{1}\}$, and $\sigma_{zx} \in \{0.1, 0.5\}$. The sizes of the two blocks are considered as $(n_1, n_2) \in \{(9.99 * 10^5, 10^3), (9 * 10^5, 10^5), (10^5, 9 * 10^5), (10^3, 9.99 * 10^5)\}$. We set $(\sigma^2, \sigma_u^2) = (1, 1), (1, 2)$ or $(3, 2)$, $\alpha_0 = 1, \boldsymbol{\alpha} = (0.5, 0.5)'$, and two different scenarios of $\boldsymbol{\beta}$, $\boldsymbol{\beta} = (0.2, 0.2)'$ and $\boldsymbol{\beta} = (0.1, 0.1, 0.1, 0.1)'$.

Table 1: $10^2 \times$ MSE when $\boldsymbol{\beta} = (0.2, 0.2)'$. CCUE is the cumulative coefficient updated estimator, EWOF is the estimator calculated without the first block, NCCUE is the no-correction cumulative updated estimator, and FULL is the estimator from full data.

	$(n_1, n_2) =$	$(9.99 * 10^5, 10^3)$		$(9 * 10^5, 10^5)$		$(10^5, 9 * 10^5)$		$(10^3, 9.99 * 10^5)$	
	$\sigma_{zx} =$	0.1	0.5	0.1	0.5	0.1	0.5	0.1	0.5
$\boldsymbol{\mu} = \mathbf{0}$									
$\sigma^2 = 1, \sigma_u^2 = 1$	CCUE	0.0296	0.1058	0.0010	0.0023	0.0006	0.0008	0.0005	0.0007
	EWOF	0.5116	0.7360	0.0053	0.0074	0.0006	0.0008	0.0005	0.0007
	NCCUE	2.1340	3.7389	1.9213	3.4075	0.0732	0.1582	0.0005	0.0007
	FULL	0.0005	0.0007	0.0005	0.0007	0.0005	0.0007	0.0005	0.0007
$\sigma^2 = 1, \sigma_u^2 = 2$	CCUE	0.0280	0.1027	0.0012	0.0031	0.0006	0.0009	0.0005	0.0007
	EWOF	0.5070	0.7619	0.0050	0.0074	0.0006	0.0008	0.0005	0.0007
	NCCUE	3.7542	6.0741	3.5000	5.7249	0.2430	0.5047	0.0005	0.0008
	FULL	0.0005	0.0008	0.0005	0.0007	0.0005	0.0007	0.0005	0.0007
$\sigma^2 = 3, \sigma_u^2 = 2$	CCUE	0.0293	0.1068	0.0029	0.0068	0.0017	0.0026	0.0014	0.0022
	ECUR	1.5468	2.3193	0.0156	0.0226	0.0017	0.0024	0.0014	0.0022
	NCCUE	3.7513	6.0749	3.4997	5.7263	0.2438	0.5047	0.0015	0.0022
	FULL	0.0015	0.0023	0.0015	0.0022	0.0016	0.0022	0.0014	0.0022
$\boldsymbol{\mu} = \mathbf{1}$									
$\sigma^2 = 1, \sigma_u^2 = 1$	CCUE	0.2217	0.4325	0.0034	0.0053	0.0010	0.0010	0.0008	0.0008
	EWOF	0.8404	0.9128	0.0081	0.0090	0.0010	0.0010	0.0008	0.0008
	NCCUE	4.9007	4.3134	4.4106	3.9306	0.1676	0.1824	0.0008	0.0009
	FULL	0.0008	0.0009	0.0008	0.0009	0.0009	0.0009	0.0008	0.0008
$\sigma^2 = 1, \sigma_u^2 = 2$	CCUE	0.2133	0.4730	0.0038	0.0057	0.0009	0.0011	0.0008	0.0009
	EWOF	0.8308	0.8830	0.0082	0.0095	0.0009	0.0010	0.0008	0.0009
	NCCUE	8.6219	7.0102	8.0351	6.6053	0.5574	0.5838	0.0009	0.0010
	FULL	0.0008	0.0009	0.0008	0.0009	0.0008	0.0009	0.0008	0.0009
$\sigma^2 = 3, \sigma_u^2 = 2$	CCUE	0.2322	0.4999	0.0067	0.0100	0.0026	0.0033	0.0025	0.0027
	EWOF	2.4457	2.7161	0.0233	0.0269	0.0026	0.0030	0.0025	0.0027
	NCCUE	8.6151	7.0155	8.0291	6.6034	0.5581	0.5835	0.0026	0.0028
	FULL	0.0024	0.0027	0.0025	0.0027	0.0024	0.0027	0.0025	0.0027

To evaluate the performance of our proposed method, four different estimators are considered for the comparison: the proposed cumulative coefficient updated estimator (CCUE), an estimator calculated without the first block (EWOF), a no-correction cumulative updated estimator (NCCUE), and the estimator from full data without any measurement error (FULL). The NCCUE are updated without the correction of the previous (biased) cumulative estimates. The EWOF are constructed with only the second block in this example. We repeat the simulation for 1000 times to calculate empirical MSEs.

The results for MSE are in Table 1 and 2 corresponding to the two scenarios for $\boldsymbol{\beta}$. First, MSEs the for FULL estimates decrease as $\boldsymbol{\mu}, \sigma_{zx}, \sigma^2$ and σ_u^2 decrease, and MSEs from other methods are

Table 2: $10^2 \times$ MSE when $\boldsymbol{\beta} = (0.1, 0.1, 0.1, 0.1)'$. CCUE is the cumulative coefficient updated estimator, EWOF is the estimator calculated without the first block, NCCUE is the no-correction cumulative updated estimator, and FULL is the estimator from full data.

	$(n_1, n_2) =$	$(9.99*10^5, 10^3)$		$(9*10^5, 10^5)$		$(10^5, 9*10^5)$		$(10^3, 9.99*10^5)$	
	$\sigma_{zx} =$	0.1	0.5	0.1	0.5	0.1	0.5	0.1	0.5
$\boldsymbol{\mu} = \mathbf{0}$									
$\sigma^2 = 1, \sigma_u^2 = 1$	CCUE	0.0282	0.1447	0.0014	0.0036	0.0008	0.0013	0.0007	0.0011
	EWOF	0.7400	1.1235	0.0072	0.0111	0.0008	0.0013	0.0007	0.0011
	NCCUE	1.1276	2.0970	1.0159	1.8947	0.0401	0.0769	0.0007	0.0011
	FULL	0.0007	0.0011	0.0007	0.0011	0.0008	0.0011	0.0007	0.0011
$\sigma^2 = 1, \sigma_u^2 = 2$	CCUE	0.0285	0.1396	0.0018	0.0049	0.0008	0.0015	0.0007	0.0011
	EWOF	0.7405	1.0719	0.0074	0.0114	0.0008	0.0013	0.0007	0.0011
	NCCUE	1.9680	3.6000	1.8344	3.3636	0.1310	0.2517	0.0008	0.0012
	FULL	0.0007	0.0011	0.0007	0.0011	0.0007	0.0011	0.0007	0.0011
$\sigma^2 = 3, \sigma_u^2 = 2$	CCUE	0.0352	0.1567	0.0061	0.0152	0.0026	0.0047	0.0022	0.0036
	EWOF	2.2117	3.5363	0.0210	0.0330	0.0025	0.0039	0.0022	0.0035
	NCCUE	2.4680	4.4767	2.3425	4.2578	0.2503	0.4769	0.0022	0.0036
	FULL	0.0022	0.0035	0.0022	0.0034	0.0022	0.0035	0.0022	0.0035
$\boldsymbol{\mu} = \mathbf{1}$									
$\sigma^2 = 1, \sigma_u^2 = 1$	CCUE	0.3786	0.6881	0.0048	0.0081	0.0013	0.0016	0.0011	0.0013
	EWOF	1.1218	1.2924	0.0112	0.0129	0.0012	0.0014	0.0011	0.0013
	NCCUE	3.9645	2.5906	3.5691	2.3398	0.1389	0.0951	0.0012	0.0013
	FULL	0.0011	0.0013	0.0012	0.0013	0.0011	0.0013	0.0011	0.0013
$\sigma^2 = 1, \sigma_u^2 = 2$	CCUE	0.4011	0.7275	0.0060	0.0092	0.0013	0.0018	0.0011	0.0013
	EWOF	1.1052	1.3161	0.0112	0.0129	0.0012	0.0015	0.0011	0.0013
	NCCUE	6.9070	4.4445	6.4397	4.1558	0.4596	0.3105	0.0012	0.0013
	FULL	0.0011	0.0013	0.0012	0.0013	0.0011	0.0013	0.0011	0.0013
$\sigma^2 = 3, \sigma_u^2 = 2$	CCUE	0.3849	0.9236	0.0124	0.0198	0.0040	0.0054	0.0033	0.0038
	EWOF	3.4223	3.9025	0.0347	0.0396	0.0037	0.0043	0.0033	0.0038
	NCCUE	8.6628	5.5295	8.2268	5.2609	0.8730	0.5894	0.0035	0.0039
	FULL	0.0033	0.0040	0.0032	0.0039	0.0033	0.0039	0.0033	0.0038

very close to those from FULL for $(n_1, n_2) = (10^3, 9.99*10^5)$. CCUE has better performance than NCCUE and EWOF in most cases. When the size of the first block is relatively big compared to the second block, the gaps of MSE between CCUE and the others (except for FULL) tend to increase. In general, the MSE is reduced more when $\boldsymbol{\mu}, \sigma_{zx}$, σ^2, and σ_u^2 are smaller. Second, the MSE for NCCUE is generally larger than that for EWOF. The updating estimator ignoring measurement error tends to lose more information on the true parameters in the process of the online updating algorithm. It is worse when the ratio of n_1 to n_2 is bigger.

We also investigate the performance of the suggested estimator in (16). Table A.1 and A.2 in Appendix E provide the result of empirical standard error (Emp SE) and average standard error (Avg SE) for α_1 and β_1 which are the first elements of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively. In general, the empirical and average standard error estimates are close to each other.

Figures 1 and 2, and Figures A.1 and A.2 in Appendix F present the bias and variance for estimation of α_1 and β_1 corresponding to the two scenarios for $\boldsymbol{\beta}$ and two values of σ_{zx}. They are plotted for varying ratios between n_1 and n_2. The panels in each figure are divided by $\boldsymbol{\mu}$ as two parts, and each column is separated by σ^2 and σ_u^2. Generally, NCCUE is more biased than CCUE and EWOF, and the gaps for the bias between CCUE and NCCUE tend to be wider in the cases $(n_1, n_2) = (9*10^5, 10^5)$ and $(10^5, 9*10^5)$ than in other cases. The variance corresponding to CCUE is smaller than EWOF in the cases $(n_1, n_2) = (9.99*10^5, 10^3)$ and $(9*10^5, 10^5)$.

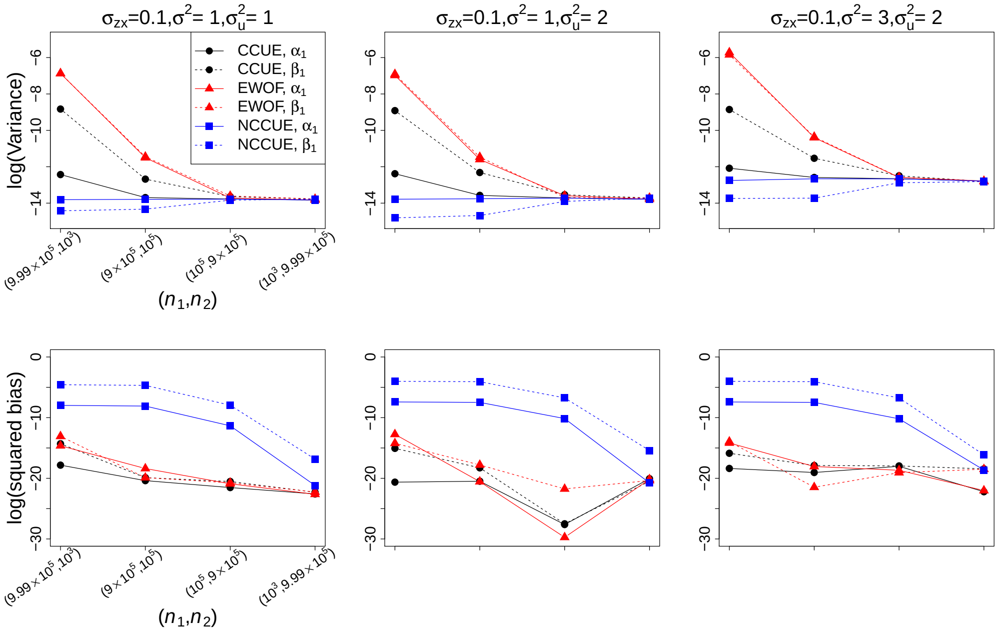
Table 3: MSE of CCUE, NCCUE, EWOF and FULL methods when $\boldsymbol{\beta} = (0.2, 0.2)', \sigma_{zx} = 0.1$ and $\boldsymbol{\mu} = \mathbf{0}$, and 4 datasets with true covariates arrive in stream. The covariates from block 2 and block 5 are not measured with error. CCUE is the cumulative coefficient updated estimator, EWOF is the estimator calculated without the first block, NCCUE is the no-correction cumulative updated estimator, and FULL is the estimator from full data.

		Block 2	Block 3	Block 4	Block 5
$\sigma^2 = 1, \sigma_u^2 = 1$	CCUE	0.0005	0.0003	0.0003	0.0002
	EWOF	0.0010	0.0005	0.0003	0.0003
	NCCUE	0.0141	0.0099	0.0073	0.0056
	FULL	0.0003	0.0003	0.0002	0.0002
$\sigma^2 = 1, \sigma_u^2 = 2$	CCUE	0.0006	0.0004	0.0003	0.0002
	EWOF	0.0010	0.0005	0.0003	0.0003
	NCCUE	0.0281	0.0216	0.0171	0.0139
	FULL	0.0003	0.0003	0.0002	0.0002
$\sigma^2 = 3, \sigma_u^2 = 2$	CCUE	0.0017	0.0011	0.0008	0.0007
	EWOF	0.0030	0.0015	0.0010	0.0008
	NCCUE	0.0286	0.0220	0.0175	0.0142
	FULL	0.0010	0.0008	0.0006	0.0005

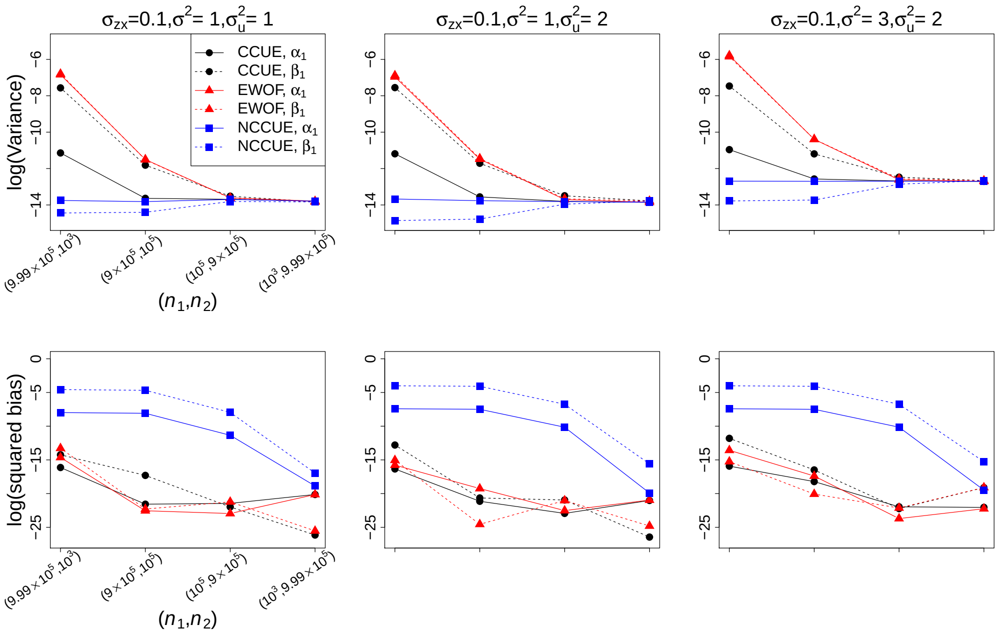
We further examine the effect of sample size of the data containing the true covariates by considering additional simulations with multiple blocks of data containing the true covariates. Table 3 shows the MSE when $\boldsymbol{\beta} = (0.2, 0.2)'$, $\boldsymbol{\mu} = \mathbf{0}$ and $\sigma_{zx} = 0.1$, and four blocks with true covariates arrive in the stream. The first block has the covariates with measurement error and the size of this block is 10,000. In blocks 2 to 5, the covariates are measured without error and the size of each of these blocks is 5,000; the simulation set-up is otherwise the same as the two-block scenario. We observe that the MSE of CCUE is smaller than the MSEs of NCCUE and EWOF. Also, as expected, the MSE of CCUE is smaller when the number of blocks with true covariates is larger. Figure A.3 and A.4 in Appendix G show the results of bias for estimating α_1, β_1, and σ^2. The estimates from CCUE perform well in terms of bias; the biases are very close to zero at all the blocks from 2 to 5, and dispersion of empirically calculated biases from all repetitions centered around zero decreases as more blocks with the true covariate arrive in the stream.

3.2. Airline On-time Data Analysis

In this section, we examine the airline on-time data obtained from the 2009 ASA Data Expo. The data contains information on the flight arrival and departure details for all commercial flights within the U.S., from October 1987 to April 2008. For the linear regression model, we used the arrival delay (in minutes) as the response and Taxi-Out time which is the elapsed time (in minutes) between departure from the airport and wheels off, Flight Time (in minutes), Security Delay (in minutes), and Elapsed Time of Flight (in minutes), as covariates with the data of 2005. For one of the covariates, Elapsed Times of Flight, there are two types; one is the Computerized Reservations Systems (CRS) Elapsed Times of Flight, and the other is Actual Elapsed Times of Flight. Similar to the simulations, we considered the case of two blocks of data. In the first block, we used the CRS Elapsed Times of Flight as the covariate from June 1, 2004 to September 29, 2005 and considered it as the covariate measured with error; the Actual Elapsed Times of Flight is used as the covariate only with data of September 30, 2005 in the second block. With the data from June 1, 2004 to September 29, 2005 including the CRS time, the linear regression model is fitted first to the first

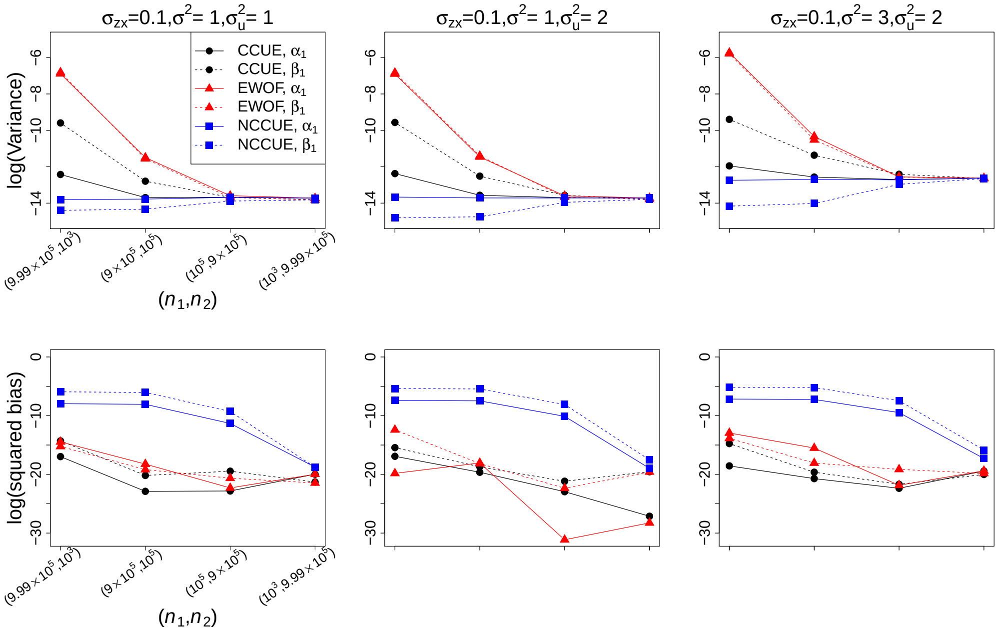


(a) $\boldsymbol{\mu} = \mathbf{0}$

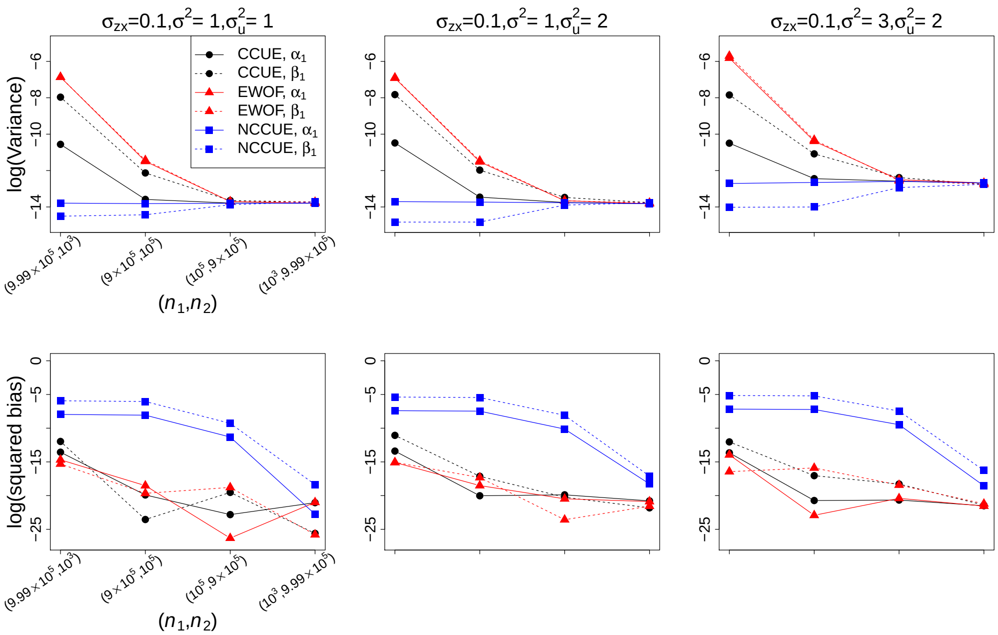


(b) $\boldsymbol{\mu} = \mathbf{1}$

Figure 1: Logarithm of variance and squared bias for estimating α_1 and β_1, the first elements of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively, when $\boldsymbol{\beta} = (0.2, 0.2)'$. CCUE is the cumulative coefficient updated estimator, EWOF is the estimator calculated without the first block, and NCCUE is the no-correction cumulative updated estimator.



(a) $\boldsymbol{\mu} = \mathbf{0}$



(b) $\boldsymbol{\mu} = \mathbf{1}$

Figure 2: Logarithm of variance and squared bias for estimating α_1 and β_1, the first elements of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ respectively when $\boldsymbol{\beta} = (0.1, 0.1, 0.1, 0.1)'$. CCUE is the cumulative coefficient updated estimator, EWOF is the estimator calculated without the first block, and NCCUE is the no-correction cumulative updated estimator.

Table 4: Estimates and standard errors of CCUE, FULL, NCCUE, ECUR, and EFIR methods for airline on-time data. CCUE is the cumulative coefficient updated estimator, NCCUE is the no-correction cumulative updated estimator, EWOF is the estimator calculated without the first block, and EFIR is the estimator from the first block of data.

	CCUE		FULL		NCCUE		EWOF		EFIR	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Intercept	-18.0462	0.2540	-16.0039	0.2391	-15.8817	0.2344	-9.8451	4.6360	-15.8835	0.2341
TaxiOut	13.3206	0.0861	12.3059	0.0493	15.1968	0.0435	9.6149	0.7594	15.2042	0.0435
FlightTime	-0.0652	0.0087	-0.0771	0.0012	0.1909	0.0011	-0.0383	0.0158	0.1914	0.0011
SecurityDelay	6.1469	0.0833	6.0793	0.0900	6.1191	0.0886	7.2822	1.8793	6.1156	0.0885
ElapsedTime	0.0599	0.0090	0.0737	0.0012	-0.2061	0.0011	0.0256	0.0164	-0.2066	0.0011

block of data and then the data with actual time on September 30, 2005 in the second block is used to update the estimates in the online-updating framework. We used five different airlines (Alaska, Hawaiian, America West, and Expressjet) and removed the observations with missing values. In addition, we exclude observations corresponding to more than 120 minutes of the arrival delay and less than 0 minutes of Flight Time. Thus, the size of first and second blocks are $n_1 = 1,023,464$ and $n_2 = 2350$, respectively. Also, two covariates, Security Delay and Taxi-Out time, are log-transformed. Table A.3 in Appendix H provides more detailed information on the Airline On-time Data.

To check autocorrelation and heteroscedasticity for error terms, we use the Durbin-Watson statistic and residual plots, respectively. The value of the Durbin-Watson statistic is 1.7429 when the first block with Actual Elapsed Time (true covariate) is used, and 1.7463 when the first block with CRS Elapsed Time (covariate with measurement error) is used. As a rule of thumb, a value less than 1 or greater than 3 is cause for concern of autocorrelation [e.g., 7]. Figure A.5 in Appendix H shows the residual plots using a random sample of 10,000 observations. Since it is hard to determine clear behavior of residuals from the full data with the large size, we did not use the full data. The Durbin-Watson statistics and the residual plots support that independence and homogeneity for errors hold approximately in the actual data.

We considered the same estimators as in the simulation study for comparison, and the results are in Table 4. For FULL estimates, Actual Elapsed Times of Flight is used for both blocks instead of CRS Elapsed Times of Flight in the first block. Furthermore, we also calculated the estimates from the first block data (EFIR) to examine the change in the estimates with and without measurement error in ElapsedTime. Overall, coefficient estimates from CCUE are similar to those from FULL. Notably, however, NCCUE and EFIR estimate for covariate ElapsedTime (the covariate measure with error) have the different sign from FULL estimate (computed from the precise measurements of ElapsedTime in both blocks 1 and 2), and the estimate from CCUE has the same sign as the FULL estimate. Also, the coefficient estimates from CCUE are closer to FULL estimates than the NCCUE and EFIR except for the estimate corresponding to SecurityDelay. Most standard errors from CCUE are larger than those from the FULL, but all of the standard errors from CCUE are much smaller than those from the EWOF.

4. Discussion

The online updating method is useful for data arriving sequentially in a stream. In this paper, we studied a method to sequentially update estimators under the situation where some covariates were initially measured with error along the data stream. At the point in which we first observe the covariates measured without error, we could consider ignoring the previous updated-estimators (as they are biased) and start the online updating process anew with the precisely measures covariates, or keep updating the estimates without the adjustment of the bias. However, these naive approaches could either lose substantial information from the historical data, or continue to be biased. Thus, we have proposed a method to correct the bias resulting from measurement error; after correction, the online-updating algorithm can be used as usual. Additionally, we have derived asymptotic results for the corrected coefficient estimators, which allows for the statistical inference.

In this study, we assume that the particular point where covariates are measured precisely is known. However, if a researcher does not know the particular point at which the covariates start to be measured precisely and continues to update without correction, the estimates could remain biased until corrected, with the degree of bias dependent on the sample sizes of data with and without the true covariates. It would be interesting to further study the case where the researcher does not know when the covariates start being measured precisely. Also, we focus on deriving theoretical results for the coefficient estimator, with SSE needed as a by-product. Development of theory for SSE, as well as non-asymptotic results for the coefficient estimator, are worthy of future research. Lastly, the online updating method under the linear errors-in-variables regression was investigated. However, it would be interesting to develop the online updating method for measurement error in more complicated, non-linear model type settings. This is another area of further investigation.

Acknowlegdements

The authors wish to thank the anonymous reviewers and associate editor for their helpful comments, which have greatly improved this manuscript. The second author was partially supported by NSF Grant 1812013.

References

[1] Carroll, R., Ruppert, D., 1996. The use and misuse of orthogonal regression in linear errors-in-variables models. The American Statistician 50, 1–6.

[2] Carroll, R.J., 1989. Covariance analysis in generalized linear measurement error models. Statistics in Medicine 8, 1075–1093.

[3] Carroll, R.J., Gail, M.H., Lubin, J.H., 1993. Case-control studies with errors in covariates. Journal of the American Statistical Association 88, 185–199.

[4] Carroll, R.J., Li, K.C., 1992. Measurement error regression with unknown link: dimension reduction and data visualization. Journal of the American Statistical Association 87, 1040–1050.

[5] Carroll, R.J., Ruppert, D., Crainiceanu, C.M., Stefanski, L.A., 2006. Measurement error in nonlinear models: a modern perspective. Chapman and Hall/CRC.

[6] Chen, X., Xie, M.g., 2014. A split-and-conquer approach for analysis of extraordinarily large data. Statistica Sinica , 1655–1684.

[7] Field, A., Miles, J., Field, Z., 2012. Discovering statistics using R. Sage publications.

[8] Fuller, W.A., 2009. Measurement error models. volume 305. John Wiley & Sons.

[9] Kleiner, A., Talwalkar, A., Sarkar, P., Jordan, M.I., 2014. A scalable bootstrap for massive data. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76, 795–816.

[10] Liang, H., Ren, H., 2005. Generalized partially linear measurement error models. Journal of Computational and Graphical Statistics 14, 237–250.

[11] Liang, H., Thurston, S.W., Ruppert, D., Apanasovich, T., Hauser, R., 2008. Additive partial linear models with measurement errors. Biometrika 95, 667–678.

[12] Lin, N., Xi, R., 2011. Aggregated estimating equation estimation. Statistics and Its Interface 4, 73–83.

[13] Ma, P., Mahoney, M.W., Yu, B., 2015. A statistical perspective on algorithmic leveraging. The Journal of Machine Learning Research 16, 861–911.

[14] Sapuppo, F., Bucolo, M., Intaglietta, M., Johnson, P.C., Fortuna, L., Arena, P., 2007. An improved instrument for real-time measurement of blood flow velocity in microvessels. IEEE Transactions on Instrumentation and Measurement 56, 2663–2671.

[15] Schifano, E.D., Wu, J., Wang, C., Yan, J., Chen, M.H., 2016. Online updating of statistical inference in the big data setting. Technometrics 58, 393–403.

[16] Song, Q., Liang, F., 2015. A split-and-merge bayesian variable selection approach for ultrahigh dimensional regression. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 77, 947–972.

[17] Stefanski, L.A., Carroll, R.J., 1987. Conditional scores and optimal scores for generalized linear measurement-error models. Biometrika 74, 703–716.

[18] Stefanski, L.A., Carroll, R.J., et al., 1985. Covariate measurement error in logistic regression. The Annals of Statistics 13, 1335–1351.

[19] Wang, C., Chen, M.H., Schifano, E., Wu, J., Yan, J., 2016a. Statistical methods and computing for big data. Statistics and its interface 9, 399.

[20] Wang, C., Chen, M.H., Wu, J., Yan, J., Zhang, Y., Schifano, E., 2018a. Online updating method with new variables for big data streams. Canadian Journal of Statistics 46, 123–146.

[21] Wang, H., Yang, M., Stufken, J., 2018b. Information-based optimal subdata selection for big data linear regression. Journal of the American Statistical Association , 1–13.

[22] Wang, H., Zhu, R., Ma, P., 2018c. Optimal subsampling for large sample logistic regression. Journal of the American Statistical Association 113, 829–844.

[23] Wang, H., Zou, G., Wan, A.T., 2013. Adaptive lasso for varying-coefficient partially linear measurement error models. Journal of Statistical Planning and Inference 143, 40–54.

[24] Wang, H., Zou, G., Wan, A.T., et al., 2012. Model averaging for varying-coefficient partially linear measurement error models. Electronic Journal of Statistics 6, 1017–1039.

[25] Wang, H.Y., Chen, X., Flournoy, N., 2016b. The focused information criterion for varying-coefficient partially linear measurement error models. Statistical Papers 57, 99–113.

[26] Wang, L., 2003. Estimation of nonlinear berkson-type measurement error models. Statistica Sinica , 1201–1210.

[27] Wang, L., et al., 2004. Estimation of nonlinear models with berkson measurement errors. The Annals of Statistics 32, 2559–2579.

[28] Wu, J., Chen, M.H., Schifano, E.D., Yan, J., 2018. Online Updating of Survival Analysis. Technical Report 18-30. University of Connecticut, Department of Statistics.

[29] Xue, Y., Wang, H., Yan, J., Schifano, E.D., 2020. An online updating approach for testing the proportional hazards assumption with streams of big survival data. Biometrics 76, 171–182.

[30] Zhang, W., Ma, H., Yang, S., 2016. An inexpensive, stable, and accurate relative humidity measurement method for challenging environments. Sensors 16, 398.

[31] Zhang, X., Wang, H., Ma, Y., Carroll, R.J., 2017. Linear model selection when covariates contain errors. Journal of the American Statistical Association 112, 1553–1561.

Supplementary Materials

Online Updating Method to Correct for Measurement Error in Big Data Streams

JooChul Lee, HaiYing Wang, and Elizabeth D. Schifano

University of Connecticut

Appendix A. Proof of Theorem 1

Write $\mathbf{T}_1 = \begin{bmatrix}\mathbf{0} & \mathbf{W}_K\end{bmatrix} = \begin{bmatrix}t_{11} & \dots & t_{1N_K}\end{bmatrix}'$ and $\mathbf{T}_2 = \begin{bmatrix}\mathbf{0} & \sqrt{\frac{N_K}{n_{K+1}}}\mathbf{X}_{K+1}\end{bmatrix} = \begin{bmatrix}t_{21} & \dots & t_{2n_{K+1}}\end{bmatrix}'$. We first discuss that $\widetilde{\mathbf{V}}_K^c/N_K$ converges in probability to $\mathbf{V}$ as N_K and $n_{K+1} \to \infty$. Since $\sum_{j=1}^{K}\sum_{i=1}^{n_j}\mathbf{z}_{ji}\mathbf{z}_{ji}'/N_K$, $\sum_{j=1}^{K}\sum_{i=1}^{n_j}\mathbf{z}_{ji}\mathbf{w}_{ji}'/N_K$, $\sum_{j=1}^{K}\sum_{i=1}^{n_j}\mathbf{w}_{ji}\mathbf{z}_{ji}'/N_K$, and $\sum_{i=1}^{n_{K+1}}\mathbf{x}_{K+1,i}\mathbf{x}_{K+1,i}'/n_{K+1}$ converge in probability to $\mathbf{E}(\mathbf{z}\mathbf{z}')$, $\mathbf{E}(\mathbf{z}\mathbf{x}')$, $\mathbf{E}(\mathbf{x}\mathbf{z}')$, and $\mathbf{E}(\mathbf{x}\mathbf{x}')$, respectively, by weak law of large numbers, we derive that $\widetilde{\mathbf{V}}_K^c/N_K = \begin{bmatrix}\sum_{j=1}^{K}\sum_{i=1}^{n_j}\mathbf{z}_{ji}\mathbf{z}_{ji}'/N_K & \sum_{j=1}^{K}\sum_{i=1}^{n_j}\mathbf{z}_{ji}\mathbf{w}_{ji}'/N_K \\ \sum_{j=1}^{K}\sum_{i=1}^{n_j}\mathbf{w}_{ji}\mathbf{z}_{ji}'/N_K & \sum_{i=1}^{n_{K+1}}\mathbf{x}_{K+1,i}\mathbf{x}_{K+1,i}'/n_{K+1}\end{bmatrix}$ converges in probability to $\mathbf{V}$. Now we prove Theorem 1. For the corrected cumulative coefficient estimator for $\boldsymbol{\theta}$ in the equation (1.5), we have

$$
\begin{aligned}
&\tilde{\boldsymbol{\theta}}_K^c - \boldsymbol{\theta} \\
&= (\widetilde{\mathbf{V}}_K^c)^{-1}\widetilde{\mathbf{V}}_K\tilde{\boldsymbol{\theta}}_K - (\widetilde{\mathbf{V}}_K^c)^{-1}(\widetilde{\mathbf{V}}_K^c)\boldsymbol{\theta} \\
&= (\widetilde{\mathbf{V}}_K^c)^{-1}\left(\sum_{j=1}^{K}\widetilde{\boldsymbol{\mathcal{X}}}_j'\mathbf{y}_j - \widetilde{\mathbf{V}}_K\boldsymbol{\theta} + \mathbf{T_1}'\mathbf{T_1}\boldsymbol{\theta} - \mathbf{T_2}'\mathbf{T_2}\boldsymbol{\theta}\right) \\
&= (\widetilde{\mathbf{V}}_K^c)^{-1}\left(\sum_{j=1}^{K}\widetilde{\boldsymbol{\mathcal{X}}}_j'\boldsymbol{\mathcal{X}}_j\boldsymbol{\theta} + \sum_{j=1}^{K}\widetilde{\boldsymbol{\mathcal{X}}}_j'\boldsymbol{\epsilon}_j - \widetilde{\mathbf{V}}_K\boldsymbol{\theta} + \mathbf{T_1}'\mathbf{T_1}\boldsymbol{\theta} - \mathbf{T_2}'\mathbf{T_2}\boldsymbol{\theta}\right) \\
&= (\widetilde{\mathbf{V}}_K^c)^{-1}\left(\sum_{j=1}^{K}\widetilde{\boldsymbol{\mathcal{X}}}_j'\begin{bmatrix}\mathbf{0} & -\mathbf{U}_j\end{bmatrix}\boldsymbol{\theta} + \sum_{j=1}^{K}\widetilde{\boldsymbol{\mathcal{X}}}_j'\boldsymbol{\epsilon}_j + \mathbf{T_1}'\mathbf{T_1}\boldsymbol{\theta} - \mathbf{T_2}'\mathbf{T_2}\boldsymbol{\theta}\right) \\
&= (\widetilde{\mathbf{V}}_K^c/N_K)^{-1}N_K^{-1}\left(\sum_{j=1}^{K}\sum_{i=1}^{n_j}\begin{bmatrix}\mathbf{z}_{ji}\\ \mathbf{w}_{ji}\end{bmatrix}\varepsilon_i + \sum_{j=1}^{K}\sum_{i=1}^{n_j}\begin{bmatrix}\mathbf{z}_{ji}\\ \mathbf{w}_{ji}\end{bmatrix}\begin{bmatrix}\mathbf{0} & -\mathbf{u}_{ji}'\end{bmatrix}\boldsymbol{\theta} + \sum_{i=1}^{N_K}t_{1i}t_{1i}'\boldsymbol{\theta} - \sum_{i=1}^{n_{K+1}}t_{2i}t_{2i}'\boldsymbol{\theta}\right) \\
&= (\widetilde{\mathbf{V}}_K^c/N_K)^{-1}(\mathbf{B}_1 - \mathbf{B}_2),
\end{aligned}
$$

where

$$\mathbf{B}_1 = N_K^{-1}\sum_{j=1}^{K}\sum_{i=1}^{n_j}\begin{bmatrix}\mathbf{z}_{ji}\varepsilon_i - \mathbf{z}_{ji}\mathbf{u}'_{ji}\boldsymbol{\beta} \\ \mathbf{w}_{ji}\varepsilon_i + (\mathbf{u}_{ji}\mathbf{x}'_{ji} + \mathbf{x}_{ji}\mathbf{x}'_{ji} - \mathbf{E}(\mathbf{x}\mathbf{x}'))\boldsymbol{\beta}\end{bmatrix} \tag{A.1}$$

and

$$\mathbf{B}_2 = N_K^{-1}\sum_{i=1}^{n_{K+1}}\begin{bmatrix}\mathbf{0} \\ (N_K/n_{K+1})(\mathbf{x}_{K+1,i}\mathbf{x}'_{K+1,i} - \mathbf{E}(\mathbf{x}\mathbf{x}'))\boldsymbol{\beta}\end{bmatrix}. \tag{A.2}$$

By Central Limit Theorem, $\mathbf{B}_1 = O_p(1/\sqrt{N_K})$ and $\mathbf{B}_2 = O_p(1/\sqrt{n_{K+1}})$.

Case 1: $n_{K+1}/N_K \to 0$.

$$\begin{aligned}\sqrt{n_{K+1}}(\tilde{\boldsymbol{\theta}}_K^c - \boldsymbol{\theta}) &= (\widetilde{\mathbf{V}}_K^c/N_K)^{-1}\sqrt{n_{K+1}}\,(\mathbf{B}_1 - \mathbf{B}_2) \\ &= (\widetilde{\mathbf{V}}_K^c/N_K)^{-1} n_{K+1}^{-1/2}\sum_{i=1}^{n_{K+1}}\begin{bmatrix}\mathbf{0} \\ (\mathbf{E}(\mathbf{x}\mathbf{x}') - \mathbf{x}_{K+1,i}\mathbf{x}'_{K+1,i})\boldsymbol{\beta}\end{bmatrix} + o_p(1) \\ &\xrightarrow{d} N(0, \mathbf{V}^{-1}\mathbf{F}_1\mathbf{V}^{-1}),\end{aligned}$$

where $\mathbf{F}_1 = \begin{bmatrix}\mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Var}(\mathbf{x}\mathbf{x}'\boldsymbol{\beta})\end{bmatrix}$. Using the Central Limit Theorem and Slutsky's Theorem, we derive the asymptotic distribution in the last step.

Case 2: $N_K/n_{K+1} \to 0$.

$$\begin{aligned}\sqrt{N_K}(\tilde{\boldsymbol{\theta}}_K^c - \boldsymbol{\theta}) &= (\widetilde{\mathbf{V}}_K^c/N_K)^{-1}\sqrt{N_K}\,(\mathbf{B}_1 - \mathbf{B}_2) \\ &= (\widetilde{\mathbf{V}}_K^c/N_K)^{-1} N_K^{-1/2}\sum_{j=1}^{K}\sum_{i=1}^{n_j}\begin{bmatrix}\mathbf{z}_{ji}\varepsilon_i - \mathbf{z}_{ji}\mathbf{u}'_{ji}\boldsymbol{\beta} \\ \mathbf{w}_{ji}\varepsilon_i + (\mathbf{u}_{ji}\mathbf{x}'_{ji} + \mathbf{x}_{ji}\mathbf{x}'_{ji} - \mathbf{E}(\mathbf{x}\mathbf{x}'))\boldsymbol{\beta}\end{bmatrix} + o_p(1) \\ &\xrightarrow{d} N(0, \mathbf{V}^{-1}\mathbf{F}_2\mathbf{V}^{-1}),\end{aligned}$$

where $\mathbf{F}_2 = \mathbf{Var}\left(\begin{bmatrix}\mathbf{z}\varepsilon_i - \mathbf{z}\mathbf{u}'\boldsymbol{\beta} \\ \mathbf{w}\varepsilon_i + (\mathbf{u}\mathbf{x}' + \mathbf{x}\mathbf{x}' - \mathbf{E}(\mathbf{x}\mathbf{x}'))\boldsymbol{\beta}\end{bmatrix}\right) = \begin{bmatrix}f_{11} & f_{12} \\ f_{21} & f_{22}\end{bmatrix}$. Central Limit Theorem and Slutsky's Theorem are applied to derive the asymptotic distribution in the last step. The elements in $\mathbf{F}_2$ can be simplified as follows.

$$\begin{aligned}f_{11} &= \mathbf{Var}\,(\mathbf{z}\varepsilon_i - \mathbf{z}\mathbf{u}'\boldsymbol{\beta}) \\ &= \mathbf{E}(\mathbf{z}\varepsilon_i^2\mathbf{z}' - \mathbf{z}\varepsilon_i\boldsymbol{\beta}'\mathbf{u}\mathbf{z}' - \mathbf{z}\mathbf{u}'\boldsymbol{\beta}\varepsilon_i\mathbf{z}' + \mathbf{z}\mathbf{u}'\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{u}\mathbf{z}') \\ &= \mathbf{E}(\mathbf{z}\mathbf{z}')\mathbf{E}(\varepsilon_i^2) - \mathbf{E}(\mathbf{z}\boldsymbol{\beta}'\mathbf{u}\mathbf{z}')\mathbf{E}(\varepsilon_i) - \mathbf{E}(\mathbf{z}\mathbf{u}'\boldsymbol{\beta}\mathbf{z}')\mathbf{E}(\varepsilon_i) + \mathbf{E}(\mathbf{z}\mathbf{z}')\mathbf{E}(\mathbf{u}'\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{u}) \\ &= \mathbf{E}(\mathbf{z}\mathbf{z}')\sigma^2 + \mathbf{E}(\mathbf{z}\mathbf{z}')\mathbf{E}(\boldsymbol{\beta}'\mathbf{u}\mathbf{u}'\boldsymbol{\beta}) \\ &= \mathbf{E}(\mathbf{z}\mathbf{z}')(\sigma^2\mathbf{I} + \boldsymbol{\beta}'\Sigma_u\boldsymbol{\beta})\end{aligned} \tag{A.3}$$

and

$$
\begin{aligned}
f_{12} &= \mathbf{E}(\mathbf{z}\varepsilon_i^2\mathbf{w}' + \mathbf{z}\varepsilon_i\boldsymbol{\beta}'(\mathbf{xx}' - \mathbf{E}(\mathbf{xx}')) + \mathbf{xu}') - \mathbf{zu}'\boldsymbol{\beta}\varepsilon_i\mathbf{w}' - \mathbf{zu}'\boldsymbol{\beta}\boldsymbol{\beta}'(\mathbf{xx}' - \mathbf{E}(\mathbf{xx}') + \mathbf{xu}')) \\
&= \mathbf{E}(\mathbf{zw}')\mathbf{E}(\varepsilon_i^2) + \mathbf{E}(\varepsilon_i)\mathbf{E}(\mathbf{z}\boldsymbol{\beta}'(\mathbf{xx}' - \mathbf{E}(\mathbf{xx}') + \mathbf{xu}') - \mathbf{zu}'\boldsymbol{\beta}\mathbf{w}') \\
&\quad -\mathbf{E}(\mathbf{zu}'\boldsymbol{\beta}\boldsymbol{\beta}'(\mathbf{xx}' - \mathbf{E}(\mathbf{xx}') + \mathbf{xu}')) \\
&= \mathbf{E}(\mathbf{zx}')\sigma^2 - \mathbf{E}(\mathbf{z})\mathbf{E}(\mathbf{u}')\boldsymbol{\beta}\boldsymbol{\beta}'(\mathbf{xx}' - \mathbf{E}(\mathbf{xx}')) - \mathbf{E}(\mathbf{zx}'\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{uu}') \\
&= \mathbf{E}(\mathbf{zx}')\sigma^2 - \mathbf{E}(\mathbf{zx}')\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{E}(\mathbf{uu}') \\
&= \mathbf{E}(\mathbf{zx}')(\sigma^2\mathbf{I} - \boldsymbol{\beta}\boldsymbol{\beta}'\Sigma_u) \\
&= f_{21}'.
\end{aligned}
$$

Also, we simplify the last element as

$$
\begin{aligned}
f_{22} &= \mathbf{E}[\mathbf{w}\varepsilon_i(\varepsilon_i\mathbf{w}' + \boldsymbol{\beta}'(\mathbf{xx}' - \mathbf{E}(\mathbf{xx}') + \mathbf{xu}')) \\
&\quad +(\mathbf{ux}' + \mathbf{xx}' - \mathbf{E}(\mathbf{xx}'))\boldsymbol{\beta}(\varepsilon_i\mathbf{w}' + \boldsymbol{\beta}'(\mathbf{xx}' - \mathbf{E}(\mathbf{xx}') + \mathbf{xu}'))] \\
&= \mathbf{E}(\mathbf{ww}')\mathbf{E}(\varepsilon_i^2) + \mathbf{E}((\mathbf{xx}' - \mathbf{E}(\mathbf{xx}') + \mathbf{ux}')\boldsymbol{\beta}\boldsymbol{\beta}'(\mathbf{xx}' - \mathbf{E}(\mathbf{xx}') + \mathbf{xu})) \\
&= (\mathbf{E}(\mathbf{xx}') + \boldsymbol{\Sigma}_u)\sigma^2 + \mathbf{E}[(\mathbf{xx}' - \mathbf{E}(\mathbf{xx}'))\boldsymbol{\beta}\boldsymbol{\beta}'(\mathbf{xx}' - \mathbf{E}(\mathbf{xx}'))] + \mathbf{E}(\mathbf{ux}'\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{xu}') \\
&= (\mathbf{E}(\mathbf{xx}') + \boldsymbol{\Sigma}_u)\sigma^2 + \mathbf{Var}(\mathbf{xx}'\boldsymbol{\beta}) + \mathbf{E}(\mathbf{uu}')\mathbf{E}(\mathbf{x}'\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{x}) \\
&= (\mathbf{E}(\mathbf{xx}') + \boldsymbol{\Sigma}_u)\sigma^2 + \mathbf{Var}(\mathbf{xx}'\boldsymbol{\beta}) + \Sigma_u\boldsymbol{\beta}'\mathbf{E}(\mathbf{xx}')\boldsymbol{\beta}.
\end{aligned}
$$

Case 3: $N_k/n_{K+1} \to h$ for some constant $0 < h < \infty$.

$$
\begin{aligned}
\sqrt{N_K}(\tilde{\boldsymbol{\theta}}_K^c - \boldsymbol{\theta}) &= (\widetilde{\mathbf{V}}_K^c/N_K)^{-1}\sqrt{N_K}\,(\mathbf{B}_1 - \mathbf{B}_2) \\
&= (\widetilde{\mathbf{V}}_K^c/N_K)^{-1}\left\{ N_K^{-1/2}\left(\sum_{j=1}^{K}\sum_{i=1}^{n_j}\begin{bmatrix}\mathbf{z}_{ji}\varepsilon_i - \mathbf{z}_{ji}\mathbf{u}_{ji}'\boldsymbol{\beta} \\ \mathbf{w}_{ji}\varepsilon_i + (\mathbf{u}_{ji}\mathbf{x}_{ji}' + \mathbf{x}_{ji}\mathbf{x}_{ji}' - \mathbf{E}(\mathbf{xx}')\boldsymbol{\beta}\end{bmatrix}\right)\right. \\
&\quad \left. -\sqrt{N_K/n_{K+1}}\,n_{K+1}^{-1/2}\left(\sum_{i=1}^{n_{K+1}}\begin{bmatrix}\mathbf{0} \\ (\mathbf{x}_{K+1,i}\mathbf{x}_{K+1,i}' - \mathbf{E}(\mathbf{xx}')\boldsymbol{\beta}\end{bmatrix}\right)\right\} \\
&\xrightarrow{d} N(0, \mathbf{V}^{-1}\mathbf{F}_3\mathbf{V}^{-1}),
\end{aligned}
$$

where $\mathbf{F}_3 = \mathbf{F}_2 + h\mathbf{F_1}$. Again, the asymptotic distribution is derived in the last step by the Central Limit Theorem and Slutsky's Theorem.

Appendix B. Derivation of the asymptotic result in Remark 1

Note that $\begin{bmatrix}\mathbf{I}_{p_1} & \mathbf{0}\end{bmatrix}(\widetilde{\mathbf{V}}_K^c/N_K)(\tilde{\boldsymbol{\theta}}_K^c - \boldsymbol{\theta}) = O_p(1/\sqrt{N_K})$. Also,

$$
\begin{aligned}
\sqrt{N_K}\begin{bmatrix}\mathbf{I}_{p_1} & \mathbf{0}\end{bmatrix}(\widetilde{\mathbf{V}}_K^c/N_K)(\tilde{\boldsymbol{\theta}}_K^c - \boldsymbol{\theta}) &= \sqrt{N_K}\begin{bmatrix}\mathbf{I}_{p_1} & \mathbf{0}\end{bmatrix}(\widetilde{\mathbf{V}}_K^c/N_K)(\widetilde{\mathbf{V}}_K^c/N_K)^{-1}(\mathbf{B}_1 - \mathbf{B}_2) \\
&= \sqrt{N_K}\begin{bmatrix}\mathbf{I}_{p_1} & \mathbf{0}\end{bmatrix}(\mathbf{B}_1 - \mathbf{B}_2) \\
&= \frac{1}{\sqrt{N_K}}\sum_{i=1}^{n_j}\left[\mathbf{z}_{ji}\varepsilon_i - \mathbf{z}_{ji}\mathbf{u}_{ji}'\boldsymbol{\beta}\right] + o_p(1) \\
&\xrightarrow{d} N(\mathbf{0}, f_{11}),
\end{aligned}
$$

where f_{11} is the same as (A.3). In the last steps, we can derive the asymptotic distribution using the Central Limit Theorem and Slutsky's Theorem.

Appendix C. Derivation of the approximation to $\mathbf{SSE}_K$

Assume that $\mathbf{u}, \epsilon$ and $(\mathbf{z}, \mathbf{x})$ are mutually independent. The second term of the right-hand side in equation (1.10) is

$$\begin{aligned}\sum_{j=1}^{K} \mathbf{A}_j' \mathbf{A}_j &= \sum_{j=1}^{K} (\tilde{\boldsymbol{\mathcal{X}}}_j \tilde{\boldsymbol{\theta}}_K - \boldsymbol{\mathcal{X}}_j \hat{\boldsymbol{\theta}}_K)' (\tilde{\boldsymbol{\mathcal{X}}}_j \tilde{\boldsymbol{\theta}}_K - \boldsymbol{\mathcal{X}}_j \hat{\boldsymbol{\theta}}_K) \\ &= \sum_{j=1}^{K} \Big(\left[\mathbf{0} \quad \mathbf{U}_j\right] \tilde{\boldsymbol{\theta}}_K + \boldsymbol{\mathcal{X}}_j \tilde{\boldsymbol{\theta}}_K - \boldsymbol{\mathcal{X}}_j \hat{\boldsymbol{\theta}}_K \Big)' \times \Big(\left[\mathbf{0} \quad \mathbf{U}_j\right] \tilde{\boldsymbol{\theta}}_K + \boldsymbol{\mathcal{X}}_j \tilde{\boldsymbol{\theta}}_K - \boldsymbol{\mathcal{X}}_j \hat{\boldsymbol{\theta}}_K \Big).\end{aligned}$$

Note that

$$\begin{aligned}\sum_{j=1}^{K} \Big(\left[\mathbf{0} \quad \mathbf{U}_j\right] \tilde{\boldsymbol{\theta}}_K \Big)' \Big(\left[\mathbf{0} \quad \mathbf{U}_j\right] \tilde{\boldsymbol{\theta}}_K \Big) &= \tilde{\mathbf{b}}_K' \sum_{j=1}^{K} \mathbf{U}_j' \mathbf{U}_j \tilde{\mathbf{b}}_K \\ &= \tilde{\mathbf{b}}_K' \sum_{j=1}^{K} \left(\sum_{i=1}^{n_j} \mathbf{w}_{ji} \mathbf{w}_{ji}' - \sum_{i=1}^{n_j} \mathbf{x}_{ji} \mathbf{x}_{ji}' \right) \tilde{\mathbf{b}}_K \\ &\quad - \tilde{\mathbf{b}}_K' \sum_{j=1}^{K} \left(\sum_{i=1}^{n_j} \mathbf{u}_{ji}' \mathbf{x}_{ji} + \sum_{i=1}^{n_j} \mathbf{x}_{ji} \mathbf{u}_{ji}' \right) \tilde{\mathbf{b}}_K\end{aligned} \tag{C.1}$$

and

$$\sum_{j=1}^{K} \Big(\left[\mathbf{0} \quad \mathbf{U}_j\right] \tilde{\boldsymbol{\theta}}_K \Big)' \Big(\boldsymbol{\mathcal{X}}_j \tilde{\boldsymbol{\theta}}_K - \boldsymbol{\mathcal{X}}_j \hat{\boldsymbol{\theta}}_K \Big) = \tilde{\mathbf{b}}_K' \sum_{j=1}^{K} \sum_{i=1}^{n_j} \mathbf{u}_{ji}' \mathbf{x}_{ji} (\tilde{\mathbf{b}}_K - \hat{\boldsymbol{\beta}}_K). \tag{C.2}$$

Since $\sum_{j=1}^{K} \sum_{i=1}^{n_j} \mathbf{u}_{ji}' \mathbf{x}_{ji} / N_K$ converges to $\mathbf{E}(\mathbf{u}'\mathbf{x}) = \mathbf{E}(\mathbf{u}')\mathbf{E}(\mathbf{x}) = \mathbf{0}$ by the weak law of large numbers, $\sum_{j=1}^{K} \sum_{i=1}^{n_j} \mathbf{u}_{ji}' \mathbf{x}_{ji}$ in equation (C.1) and (C.2) is of order $o_p(N_K)$. Then, we have

$$\begin{aligned}\sum_{j=1}^{K} \mathbf{A}_j' \mathbf{A}_j &= \sum_{j=1}^{K} (\tilde{\boldsymbol{\theta}}_K - \hat{\boldsymbol{\theta}}_K)' \boldsymbol{\mathcal{X}}_j' \boldsymbol{\mathcal{X}}_j (\tilde{\boldsymbol{\theta}}_K - \hat{\boldsymbol{\theta}}_K) \\ &\quad + \hat{\mathbf{b}}_K' \sum_{j=1}^{K} \left(\sum_{i=1}^{n_j} \mathbf{w}_{ji} \mathbf{w}_{ji}' - \sum_{i=1}^{n_j} \mathbf{x}_{ji} \mathbf{x}_{ji}' \right) \hat{\mathbf{b}}_K + o_p(N_K).\end{aligned}$$

Also, the last term of the right-hand side in equation (1.10) is

$$2\sum_{j=1}^{K}(\mathbf{y}_j - \boldsymbol{\mathcal{X}}_j\hat{\boldsymbol{\theta}}_K)'\mathbf{A}_j = 2\sum_{j=1}^{K}(\mathbf{y}_j - \boldsymbol{\mathcal{X}}_j\hat{\boldsymbol{\theta}}_K)'\left(\begin{bmatrix}\mathbf{0} & \mathbf{U}_j\end{bmatrix}\tilde{\boldsymbol{\theta}}_K + \boldsymbol{\mathcal{X}}_j\tilde{\boldsymbol{\theta}}_K - \boldsymbol{\mathcal{X}}_j\hat{\boldsymbol{\theta}}_j\right)$$
$$= 2\sum_{j=1}^{K}(\mathbf{y}_j - \boldsymbol{\mathcal{X}}_j\hat{\boldsymbol{\theta}}_k)'\mathbf{U}_j\tilde{\mathbf{b}}_K.$$

Since $N_K^{-1}\sum_{j=1}^{K}(y_{ji} - \boldsymbol{\mathcal{X}}_j\hat{\boldsymbol{\theta}}_k)'\mathbf{U}_j\tilde{\mathbf{b}}_K = N_K^{-1}\sum_{j=1}^{K}\sum_{i=1}^{n_{K+1}}\left(y_{ji} - \begin{bmatrix}\mathbf{z}_{ji} & \mathbf{x}_{ji}\end{bmatrix}\hat{\boldsymbol{\theta}}_k\right)\mathbf{u}'_{ji}\tilde{\mathbf{b}}_K$ converges to $\mathbf{E}(y_{ji}\mathbf{u}'\tilde{\mathbf{b}}_K) - \mathbf{E}\left(\mathbf{u}'\begin{bmatrix}\mathbf{z} & \mathbf{x}\end{bmatrix}\hat{\boldsymbol{\theta}}_k\mathbf{b}_K\right) = \mathbf{E}(y_{ji})\mathbf{E}(\mathbf{u}')\mathbf{E}(\tilde{\mathbf{b}}_K) - \mathbf{E}(\mathbf{u}')\mathbf{E}\left(\begin{bmatrix}\mathbf{z} & \mathbf{x}\end{bmatrix}\hat{\boldsymbol{\theta}}_k\mathbf{b}_K\right) = \mathbf{0}$ by the weak law of large numbers, $2\sum_{j=1}^{K}(\mathbf{y}_j - \boldsymbol{\mathcal{X}}_j\hat{\boldsymbol{\theta}}_K)'\mathbf{A}_j$ is of order $o_p(N_K)$.

Appendix D. Proof of Theorem 2

Note that

$$\widetilde{\mathbf{V}}^c + \boldsymbol{\mathcal{X}}'_{K+1}\boldsymbol{\mathcal{X}}_{K+1} = \begin{bmatrix}\sum_{j=1}^{K+1}\sum_{i=1}^{n_j}\mathbf{z}_{ji}\mathbf{z}'_{ji} & \sum_{j=1}^{K+1}\sum_{i=1}^{n_j}\mathbf{z}_{ji}\mathbf{x}'_{ji} + \sum_{j=1}^{K}\sum_{i=1}^{n_j}\mathbf{z}_{ji}\mathbf{u}'_{ji} \\ \sum_{j=1}^{K+1}\sum_{i=1}^{n_j}\mathbf{x}_{ji}\mathbf{z}'_{ji} + \sum_{j=1}^{K}\sum_{i=1}^{n_j}\mathbf{u}_{ji}\mathbf{z}'_{ji} & (N_{K+1}/n_{K+1})\sum_{i=1}^{n_{K+1}}\mathbf{x}_{K+1,i}\mathbf{x}'_{K+1,i}\end{bmatrix}.$$

Since $\sum_{j=1}^{K}\sum_{i=1}^{n_j}\mathbf{z}_{ji}\mathbf{u}'_{ji} = o_p(N_K)$, $N_{K+1}^{-1}\left(\sum_{j=1}^{K+1}\sum_{i=1}^{n_j}\mathbf{z}_{ji}\mathbf{x}'_{ji} + \sum_{j=1}^{K}\sum_{i=1}^{n_j}\mathbf{z}_{ji}\mathbf{u}'_{ji}\right)$ and $N_{K+1}^{-1}\left(\sum_{j=1}^{K+1}\sum_{i=1}^{n_j}\mathbf{x}_{ji}\mathbf{z}'_{ji} + \sum_{j=1}^{K}\sum_{i=1}^{n_j}\mathbf{u}_{ji}\mathbf{z}'_{ji}\right)$ converge in probability to $\mathbf{E}(\mathbf{z}\mathbf{x}')$ and $\mathbf{E}(\mathbf{x}\mathbf{z}')$, respectively, by the weak law of large numbers as $N_K, n_{K+1} \to \infty$. Also, $\sum_{j=1}^{K+1}\sum_{i=1}^{n_j}\mathbf{z}_{ji}\mathbf{z}'_{ji}/N_{K+1}$ and $\sum_{i=1}^{n_{K+1}}\mathbf{x}_{K+1,i}\mathbf{x}'_{K+1,i}/n_{K+1}$ converge in probability to $\mathbf{E}(\mathbf{z}\mathbf{z}')$ and $\mathbf{E}(\mathbf{x}\mathbf{x}')$, respectively. Thus, $(\widetilde{\mathbf{V}}_K^c + \boldsymbol{\mathcal{X}}'_{K+1}\boldsymbol{\mathcal{X}}_{K+1})/N_{K+1}$ converges in probability to $\mathbf{V}$.

Denote $\Delta = (\widetilde{\mathbf{V}}_K^c + \boldsymbol{\mathcal{X}}'_{K+1}\boldsymbol{\mathcal{X}}_{K+1})/N_{K+1}$. For the cumulative coefficient estimator in the equation

(1.12), we have

$$
\begin{aligned}
&\hat{\boldsymbol{\theta}}^c_{K+1} - \boldsymbol{\theta} \\
&= \Delta^{-1} N_{K+1}^{-1} (\widetilde{\mathbf{V}}_K \tilde{\boldsymbol{\theta}}_K + \boldsymbol{\mathcal{X}}'_{K+1} \boldsymbol{\mathcal{X}}_{K+1} \hat{\boldsymbol{\vartheta}}_{n_{K+1},K+1} - (\widetilde{\mathbf{V}}^c_K + \boldsymbol{\mathcal{X}}'_{K+1} \boldsymbol{\mathcal{X}}_{K+1}) \boldsymbol{\theta}) \\
&= \Delta^{-1} N_{K+1}^{-1} \left(\sum_{j=1}^{K} \widetilde{\boldsymbol{\mathcal{X}}}'_j \mathbf{y}_j - \widetilde{\mathbf{V}}_K \boldsymbol{\theta} + \mathbf{T_1}'\mathbf{T_1}\boldsymbol{\theta} - \mathbf{T_2}'\mathbf{T_2}\boldsymbol{\theta} + \boldsymbol{\mathcal{X}}'_{K+1} \mathbf{y}_{K+1} - \boldsymbol{\mathcal{X}}'_{K+1} \boldsymbol{\mathcal{X}}_{K+1} \boldsymbol{\theta} \right) \\
&= \Delta^{-1} N_{K+1}^{-1} \left(\sum_{j=1}^{K} \widetilde{\boldsymbol{\mathcal{X}}}'_j \boldsymbol{\mathcal{X}}_j \boldsymbol{\theta} + \sum_{j=1}^{K} \widetilde{\boldsymbol{\mathcal{X}}}'_j \boldsymbol{\epsilon}_j - \widetilde{\mathbf{V}}_K \boldsymbol{\theta} + \mathbf{T_1}'\mathbf{T_1}\boldsymbol{\theta} - \mathbf{T_2}'\mathbf{T_2}\boldsymbol{\theta} + \boldsymbol{\mathcal{X}}'_{K+1} \boldsymbol{\epsilon}_{K+1} \right) \\
&= \Delta^{-1} N_{K+1}^{-1} \left(\sum_{j=1}^{K} \widetilde{\boldsymbol{\mathcal{X}}}'_j \begin{bmatrix} \mathbf{0} & -\mathbf{U}_j \end{bmatrix} \boldsymbol{\theta} + \sum_{j=1}^{K} \widetilde{\boldsymbol{\mathcal{X}}}'_j \boldsymbol{\epsilon}_j + \mathbf{T_1}'\mathbf{T_1}\boldsymbol{\theta} - \mathbf{T_2}'\mathbf{T_2}\boldsymbol{\theta} + \boldsymbol{\mathcal{X}}'_{K+1} \boldsymbol{\epsilon}_{K+1} \right) \\
&= \Delta^{-1} \left((N_K/N_{K+1})(\mathbf{B}_1 - \mathbf{B}_2) + (n_{K+1}/N_{K+1})\mathbf{B}_3 \right),
\end{aligned}
$$

where $\mathbf{B}_1$ and $\mathbf{B}_2$ are the same as (A.1) and (A.2), respectively, and $\mathbf{B}_3 = n_{K+1}^{-1} \sum_{i=1}^{n_{K+1}} \begin{bmatrix} \mathbf{z}_{K+1,i}\varepsilon_i \\ \mathbf{x}_{K+1,i}\varepsilon_i \end{bmatrix}$. By Central Limit Theorem, $\mathbf{B}_3$ is of order $O_p(1/\sqrt{n_{K+1}})$. Like the cases 1, 2, and 3 in Theorem 1, we use the Central Limit Theorem and Slutsky's Theorem in the last steps of each case to derive the asymptotic normal distribution. Note that $\mathbf{B}_1 = O_p(1/\sqrt{N_K})$, and $\mathbf{B}_2 = O_p(1/\sqrt{n_{K+1}})$.

Case 1: $n_{K+1}/N_K \to 0$. Since $\sqrt{n_{K+1}N_K}/N_{K+1} \to 0$ as $n_{K+1}/N_K \to 0$, $\left(\sqrt{n_{K+1}}N_K/N_{K+1}\right)\mathbf{B}_1 = o_p(1)$ and $\left(\sqrt{n_{K+1}}n_{K+1}/N_{K+1}\right)\mathbf{B}_3 = o_p(1)$. Note that $N_K/N_{K+1} \to 1$ as $n_{K+1}/N_K \to 0$.

$$
\begin{aligned}
\sqrt{n_{K+1}}(\hat{\boldsymbol{\theta}}^c_{K+1} - \boldsymbol{\theta}) &= \Delta^{-1} \left[\left(\sqrt{n_{K+1}}N_K/N_{K+1}\right)(\mathbf{B}_1 - \mathbf{B}_2) + \left(\sqrt{n_{K+1}}n_{K+1}/N_{K+1}\right)\mathbf{B}_3 \right] \\
&= \Delta^{-1} \left(N_K/N_{K+1}\right) \sqrt{n_{K+1}}\mathbf{B}_2 + o_p(1) \\
&\xrightarrow{d} N(0, \mathbf{V}^{-1}\mathbf{F}_1\mathbf{V}^{-1}).
\end{aligned}
$$

Case 2: $N_K/n_{K+1} \to 0$. Since $\sqrt{N_K/N_{K+1}} \to 0$ as $N_K/n_{K+1} \to 0$, $\left(N_K/\sqrt{N_{K+1}}\right)\mathbf{B}_1 = o_p(1)$ and $\left(N_K/\sqrt{N_{K+1}}\right)\mathbf{B}_3 = o_p(1)$. Note that $\sqrt{n_{K+1}/N_{K+1}} \to 1$ as $N_K/n_{K+1} \to 0$.

$$
\begin{aligned}
\sqrt{N_{K+1}}(\hat{\boldsymbol{\theta}}^c_{K+1} - \boldsymbol{\theta}) &= \Delta^{-1} \left(\left(N_K/\sqrt{N_{K+1}}\right)(\mathbf{B}_1 - \mathbf{B}_2) + \left(n_{K+1}/\sqrt{N_{K+1}}\right)\mathbf{B}_3 \right) \\
&= \Delta^{-1} \left(n_{K+1}/\sqrt{N_{K+1}}\right)\mathbf{B}_3 + o_p(1) \\
&\xrightarrow{d} N(0, \mathbf{V}^{-1}\sigma^2).
\end{aligned}
$$

Case 3: $N_K/n_{K+1} \to h$ for some constant $0 < h < \infty$.

$$\begin{aligned}
\sqrt{N_{K+1}}(\hat{\boldsymbol{\theta}}^c_{K+1} - \boldsymbol{\theta}) &= \Delta^{-1}\left[\left(N_K/\sqrt{N_{K+1}}\right)(\mathbf{B}_1 - \mathbf{B}_2) + \left(n_{K+1}/\sqrt{N_{K+1}}\right)\mathbf{B}_3\right] \\
&= \Delta^{-1}\left(\sqrt{\frac{N_K}{N_{K+1}}}\frac{1}{\sqrt{N_K}}\sum_{j=1}^{K}\sum_{i=1}^{n_j}\begin{bmatrix}\mathbf{z}_{ji}\varepsilon_i - \mathbf{z}_{ji}\mathbf{u}'_{ji}\boldsymbol{\beta} \\ \mathbf{w}_{ji}\varepsilon_i + (\mathbf{u}_{ji}\mathbf{x}'_{ji} + \mathbf{x}_{ji}\mathbf{x}'_{ji} - \mathbf{E}(\mathbf{x}\mathbf{x}'))\boldsymbol{\beta}\end{bmatrix}\right. \\
&\quad \left. + \sqrt{\frac{n_{K+1}}{N_{K+1}}}\frac{1}{\sqrt{n_{K+1}}}\sum_{i=1}^{n_{K+1}}\begin{bmatrix}\mathbf{z}_{K+1,i}\varepsilon_i \\ \frac{N_K}{n_{K+1}}(\mathbf{x}_{ji}\mathbf{x}'_{ji} - \mathbf{E}(\mathbf{x}\mathbf{x}'))\boldsymbol{\beta} + \mathbf{x}_{K+1,i}\varepsilon_i\end{bmatrix}\right) \\
&\xrightarrow{d} N(0, \mathbf{V}^{-1}\mathbf{F}_4\mathbf{V}^{-1}),
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{F}_4 &= \frac{h}{h+1}\mathbf{Var}\left(\begin{bmatrix}\mathbf{z}\varepsilon_i - \mathbf{z}\mathbf{u}'\boldsymbol{\beta} \\ \mathbf{w}\varepsilon_i + (\mathbf{u}\mathbf{x}' + \mathbf{x}\mathbf{x}' - \mathbf{E}(\mathbf{x}\mathbf{x}'))\boldsymbol{\beta}\end{bmatrix}\right) \\
&\quad + \frac{1}{h+1}\left(\mathbf{Var}\left(\begin{bmatrix}\mathbf{0} \\ h(\mathbf{x}\mathbf{x}' - \mathbf{E}(\mathbf{x}\mathbf{x}'))\boldsymbol{\beta}\end{bmatrix}\right) + \mathbf{Var}\left(\begin{bmatrix}\mathbf{z}\varepsilon_i \\ \mathbf{x}\varepsilon_i\end{bmatrix}\right)\right) \\
&= \frac{h}{h+1}\mathbf{F_2} + \frac{1}{h+1}(h^2\mathbf{F_1} + \mathbf{V}\sigma^2).
\end{aligned}$$

Appendix E. Table A.1 and A.2 for empirical and average standard error

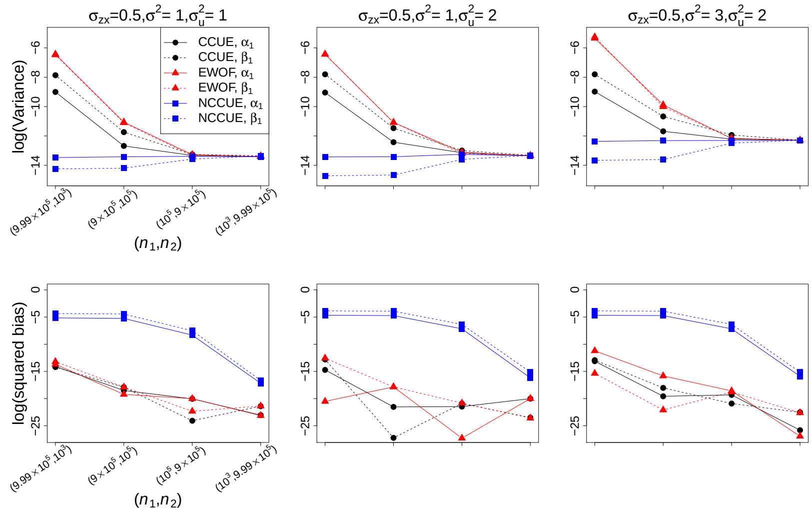
Table A.1: Empirical standard error (Emp SE) and average standard error (Avg SE) for α_1 and β_1 when $\boldsymbol{\beta} = (0.2, 0.2)'$

	$(n_1, n_2) =$	$(9.99 * 10^5, 10^3)$		$(9 * 10^5, 10^5)$		$(10^5, 9 * 10^5)$		$(10^3, 9.99 * 10^5)$	
	$\sigma_{zx} =$	0.1	0.5	0.1	0.5	0.1	0.5	0.1	0.5
$\boldsymbol{\mu} = \mathbf{0}$									
$\sigma^2 = 1, \sigma_u^2 = 1$	Emp SE(α_1)	0.0020	0.0111	0.0011	0.0018	0.0010	0.0013	0.0010	0.0012
	Avg SE(α_1)	0.0020	0.0109	0.0011	0.0018	0.0010	0.0013	0.0010	0.0013
	Emp SE(β_1)	0.0121	0.0196	0.0018	0.0028	0.0011	0.0013	0.0010	0.0013
	Avg SE(β_1)	0.0118	0.0197	0.0018	0.0028	0.0011	0.0014	0.0010	0.0013
$\sigma^2 = 1, \sigma_u^2 = 2$	Emp SE(α_1)	0.0020	0.0109	0.0011	0.0020	0.0011	0.0014	0.0010	0.0013
	Avg SE(α_1)	0.0020	0.0111	0.0011	0.0020	0.0010	0.0013	0.0010	0.0013
	Emp SE(β_1)	0.0116	0.0202	0.0021	0.0032	0.0011	0.0015	0.0011	0.0013
	Avg SE(β_1)	0.0119	0.0201	0.0021	0.0032	0.0011	0.0015	0.0010	0.0013
$\sigma^2 = 3, \sigma_u^2 = 2$	Emp SE(α_1)	0.0024	0.0112	0.0018	0.0029	0.0018	0.0022	0.0017	0.0021
	Avg SE(α_1)	0.0025	0.0115	0.0018	0.0029	0.0018	0.0023	0.0018	0.0022
	Emp SE(β_1)	0.0119	0.0202	0.0031	0.0048	0.0019	0.0026	0.0017	0.0021
	Avg SE(β_1)	0.0123	0.0208	0.0032	0.0048	0.0019	0.0025	0.0018	0.0022
$\boldsymbol{\mu} = \mathbf{1}$									
$\sigma^2 = 1, \sigma_u^2 = 1$	Emp SE(α_1)	0.0038	0.0221	0.0011	0.0025	0.0011	0.0013	0.0010	0.0012
	Avg SE(α_1)	0.0038	0.0211	0.0011	0.0025	0.0010	0.0013	0.0010	0.0013
	Emp SE(β_1)	0.0228	0.0370	0.0027	0.0041	0.0012	0.0014	0.0010	0.0012
	Avg SE(β_1)	0.0235	0.0359	0.0027	0.0040	0.0011	0.0015	0.0010	0.0013
$\sigma^2 = 1, \sigma_u^2 = 2$	Emp SE(α_1)	0.0037	0.0230	0.0011	0.0025	0.0010	0.0013	0.0010	0.0013
	Avg SE(α_1)	0.0038	0.0215	0.0012	0.0026	0.0010	0.0014	0.0010	0.0013
	Emp SE(β_1)	0.0229	0.0394	0.0029	0.0043	0.0012	0.0015	0.0010	0.0013
	Avg SE(β_1)	0.0234	0.0366	0.0029	0.0044	0.0012	0.0016	0.0010	0.0013
$\sigma^2 = 3, \sigma_u^2 = 2$	Emp SE(α_1)	0.0042	0.0236	0.0019	0.0033	0.0018	0.0023	0.0017	0.0022
	Avg SE(α_1)	0.0042	0.0220	0.0019	0.0034	0.0018	0.0023	0.0018	0.0022
	Emp SE(β_1)	0.0239	0.0391	0.0037	0.0056	0.0020	0.0026	0.0018	0.0022
	Avg SE(β_1)	0.0241	0.0373	0.0038	0.0057	0.0020	0.0026	0.0018	0.0022

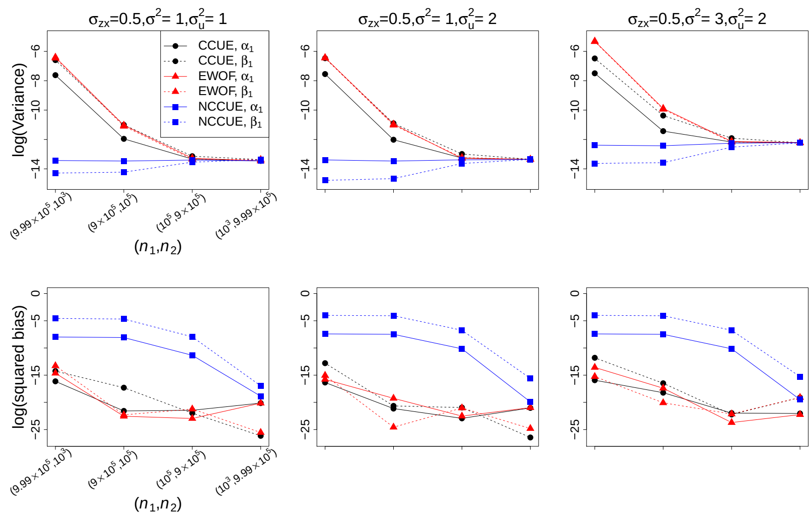
Table A.2: Empirical standard error (Emp SE) and average standard error (Avg SE) for α_1 and β_1 when $\boldsymbol{\beta} = (0.1, 0.1, 0.1, 0.1)'$

	$(n_1, n_2) =$	$(9.99 * 10^5, 10^3)$		$(9 * 10^5, 10^5)$		$(10^5, 9 * 10^5)$		$(10^3, 9.99 * 10^5)$	
	$\sigma_{zx} =$	0.1	0.5	0.1	0.5	0.1	0.5	0.1	0.5
$\boldsymbol{\mu = 0}$									
$\sigma^2 = 1, \sigma_u^2 = 1$	Emp SE(α_1)	0.0020	0.0135	0.0011	0.0020	0.0011	0.0013	0.0010	0.0013
	Avg SE(α_1)	0.0021	0.0133	0.0011	0.0019	0.0010	0.0013	0.0010	0.0013
	Emp SE(β_1)	0.0083	0.0165	0.0017	0.0026	0.0011	0.0014	0.0010	0.0014
	Avg SE(β_1)	0.0086	0.0162	0.0016	0.0026	0.0011	0.0014	0.0010	0.0013
$\sigma^2 = 1, \sigma_u^2 = 2$	Emp SE(α_1)	0.0021	0.0132	0.0011	0.0021	0.0011	0.0014	0.0011	0.0013
	Avg SE(α_1)	0.0021	0.0132	0.0011	0.0021	0.0010	0.0014	0.0010	0.0013
	Emp SE(β_1)	0.0084	0.0160	0.0019	0.0031	0.0011	0.0015	0.0010	0.0013
	Avg SE(β_1)	0.0086	0.0163	0.0019	0.0031	0.0011	0.0015	0.0010	0.0013
$\sigma^2 = 3, \sigma_u^2 = 2$	Emp SE(α_1)	0.0025	0.0136	0.0019	0.0032	0.0017	0.0024	0.0018	0.0023
	Avg SE(α_1)	0.0026	0.0141	0.0019	0.0033	0.0018	0.0024	0.0018	0.0023
	Emp SE(β_1)	0.0091	0.0168	0.0034	0.0057	0.0020	0.0029	0.0018	0.0023
	Avg SE(β_1)	0.0097	0.0179	0.0035	0.0057	0.0020	0.0028	0.0018	0.0023
$\boldsymbol{\mu = 1}$									
$\sigma^2 = 1, \sigma_u^2 = 1$	Emp SE(α_1)	0.0051	0.0303	0.0011	0.0031	0.0010	0.0013	0.0010	0.0013
	Avg SE(α_1)	0.0050	0.0285	0.0012	0.0030	0.0010	0.0014	0.0010	0.0013
	Emp SE(β_1)	0.0186	0.0297	0.0023	0.0036	0.0011	0.0014	0.0010	0.0014
	Avg SE(β_1)	0.0191	0.0297	0.0023	0.0035	0.0011	0.0015	0.0010	0.0013
$\sigma^2 = 1, \sigma_u^2 = 2$	Emp SE(α_1)	0.0053	0.0312	0.0012	0.0031	0.0010	0.0014	0.0010	0.0013
	Avg SE(α_1)	0.0050	0.0285	0.0012	0.0031	0.0010	0.0014	0.0010	0.0013
	Emp SE(β_1)	0.0200	0.0311	0.0025	0.0040	0.0012	0.0016	0.0010	0.0012
	Avg SE(β_1)	0.0192	0.0300	0.0026	0.0040	0.0012	0.0016	0.0010	0.0013
$\sigma^2 = 3, \sigma_u^2 = 2$	Emp SE(α_1)	0.0053	0.0357	0.0020	0.0039	0.0019	0.0025	0.0018	0.0023
	Avg SE(α_1)	0.0053	0.0300	0.0019	0.0040	0.0018	0.0024	0.0018	0.0023
	Emp SE(β_1)	0.0198	0.0338	0.0039	0.0060	0.0020	0.0029	0.0017	0.0023
	Avg SE(β_1)	0.0199	0.0315	0.0039	0.0062	0.0020	0.0029	0.0018	0.0023

Appendix F. Figure A.1 and Figure A.2 for logarithm of variance and squared bias for estimating α_1 and β_1

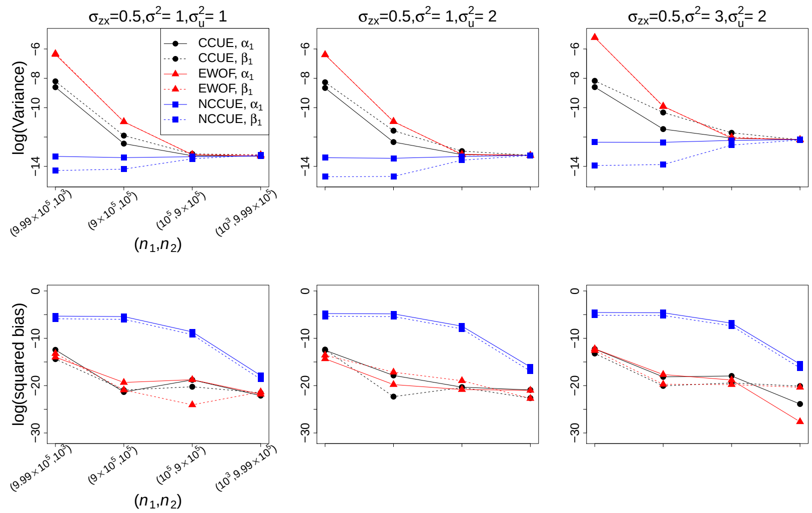


(a) $\boldsymbol{\mu} = \mathbf{0}$

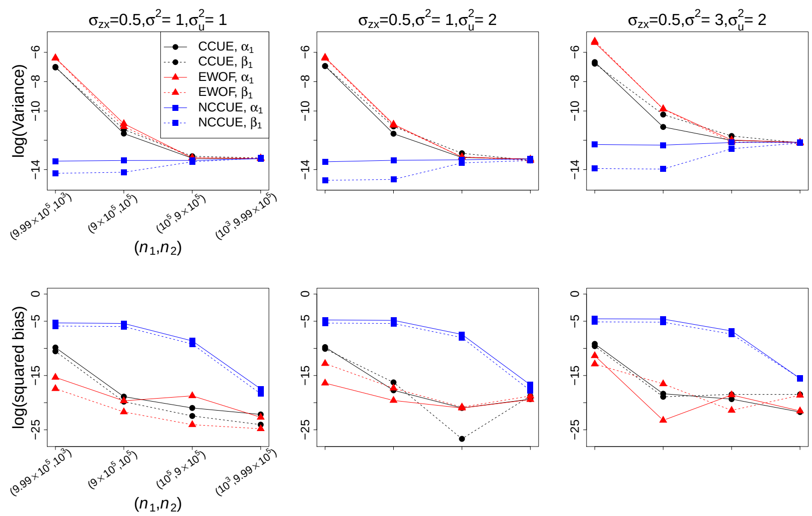


(b) $\boldsymbol{\mu} = \mathbf{1}$

Figure A.1: Logarithm of variance and squared bias for estimating α_1 and β_1, the first elements of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively, when $\boldsymbol{\beta} = (0.2, 0.2)'$. CCUE is the cumulative coefficient updated estimator, EWOF is the estimator calculated without the first block, and NCCUE is the no-correction cumulative updated estimator.



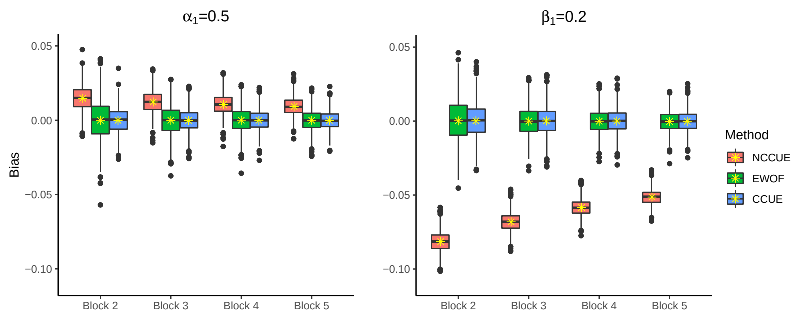
(a) $\boldsymbol{\mu} = \mathbf{0}$



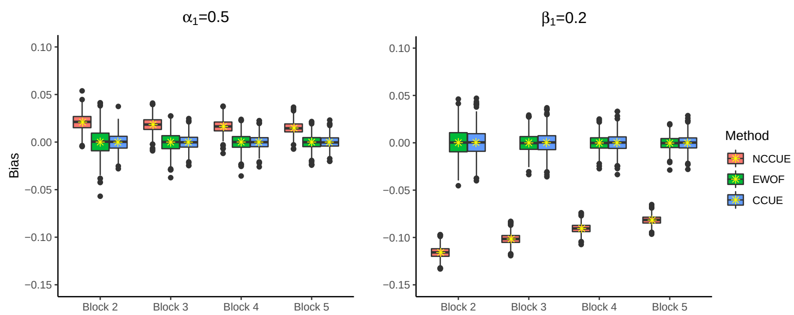
(b) $\boldsymbol{\mu} = \mathbf{1}$

Figure A.2: Logarithm of variance and squared bias for estimating α_1 and β_1, the first elements of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ respectively when $\boldsymbol{\beta} = (0.1, 0.1, 0.1, 0.1)'$. CCUE is the cumulative coefficient updated estimator, EWOF is the estimator calculated without the first block, and NCCUE is the no-correction cumulative updated estimator.

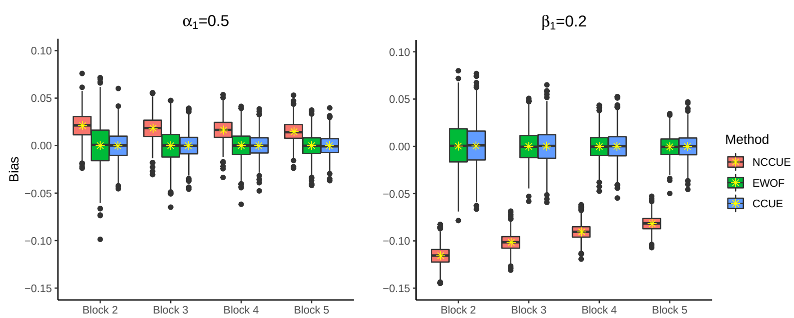
Appendix G. Figure A.3 and Figure A.4 for boxplots of bias for estimating α_1, β_1, and σ^2



(a) $\sigma^2 = 1, \sigma_u^2 = 1$



(b) $\sigma^2 = 1, \sigma_u^2 = 2$



(c) $\sigma^2 = 3, \sigma_u^2 = 2$

Figure A.3: Boxplots of bias for estimating α_1 and β_1, the first elements of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively, when $\boldsymbol{\beta} = (0.2, 0.2)'$, $\boldsymbol{\mu} = \mathbf{0}$, $\sigma_{zx} = 0.1$, and 4 datasets with true covariates arrive in stream. Note that yellow asterisk in the boxplot indicates mean. CCUE is the cumulative coefficient updated estimator, EWOF is the estimator updated without the first block, and NCCUE is the no-correction cumulative updated estimator.

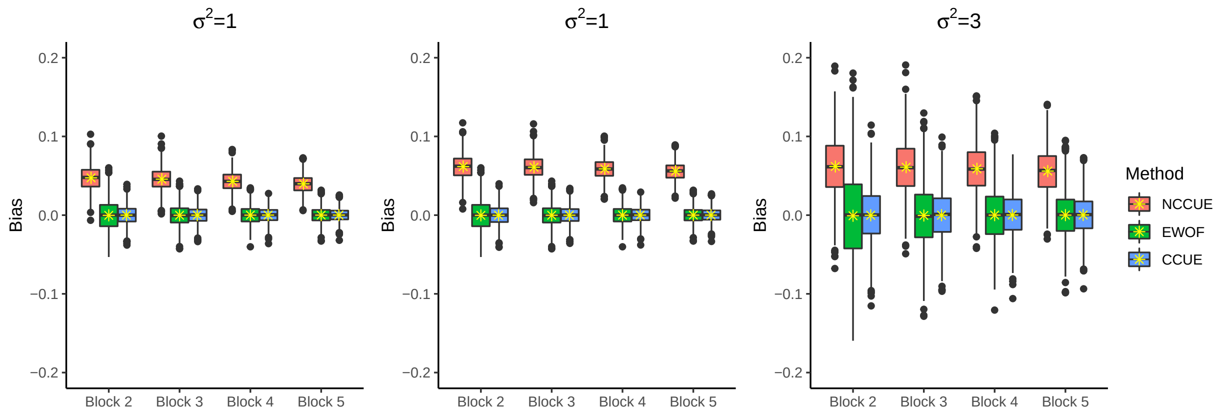


Figure A.4: Boxplots of bias for estimating σ^2 when $\boldsymbol{\beta} = (0.2, 0.2)'$, $\boldsymbol{\mu} = \mathbf{0}$, $\sigma_{zx} = 0.1$, $\sigma_u^2 = 1$ (left plot), 2 (middle and right plots), and 4 datasets with true covariates arrive in streams. Note that yellow asterisk in the boxplot indicates mean. CCUE is the cumulative coefficient updated estimator, EWOF is the estimator calculated without the first block, and NCCUE is the no-correction cumulative updated estimator.

Appendix H. Figure A.5 for heteroscedasticity and Table A.3 for the detailed information in Airline on-time data

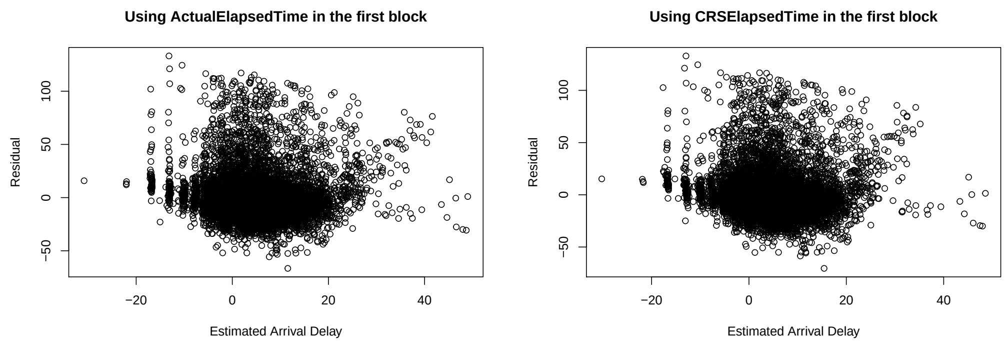


Figure A.5: Residual plots with 10,000 random sample in the Airline on-time data. 'Actual elapsed time' variable is used in the first block for the left plot, and 'CRS elapsed time' variable is used in the first block for the right plot.

Table A.3: First five observations in Airline on-time data

	Arrival Delay[1]	Taxi-Out Time[2]	Flight Time[3]	Security Delay[4]	CRS Elapsed Time[5]	Actual Elapsed Time[6]
obs 1	10	36	85	0	108	124
obs 2	-14	12	52	0	80	68
obs 3	50	13	88	0	65	110
obs 4	2	24	90	0	114	119
obs 5	21	29	46	0	74	78

[1] Difference in minutes between scheduled and actual arrival time
[2] Elapsed time in minutes between departure from the origin airport gate and wheels off
[3] The total time in minutes for an aircraft is in the air from wheels-off at the origin airport to wheels-down at the destination airport
[4] Delayed time in minutes caused by evacuation of a terminal or concourse, re-boarding of aircraft because of security breach, inoperative screening equipment and/or long lines
[5] Computer Reservation System time in minutes computed from gate departure time to gate arrival time
[6] Actual time in minutes computed from gate departure time to gate arrival time