MONASH BUSINESS SCHOOL

ISSN 1440-771X

Department of Econometrics and Business Statistics

http://business.monash.edu/econometrics-and-business-statistics/research/publications

Variable Selection for a Categorical Varying-Coefficient Model with Identifications for Determinants of Body Mass Index

Jiti Gao, Bin Peng, Zhao Ren and Xiaohui Zhang

October 2015

Working Paper 21/15

Variable Selection for a Categorical Varying-Coefficient Model with Identifications for Determinants of Body Mass Index

Jiti Gao*, Bin Peng†, Zhao Ren§ and Xiaohui Zhang‡

*Monash University, †University of Technology Sydney, §University of Pittsburgh and ‡Murdoch University

Abstract

In this paper, we propose a variable selection procedure based on the shrinkage estimation technique for a categorical varying-coefficient model. We apply the method to identify the relevant determinants for body mass index (BMI) from a large amount of potential factors proposed in the multidisciplinary literature, using data from the 2013 National Health Interview Survey in the United States. We quantify the varying impacts of the relevant determinants of BMI across demographic groups.

Keywords: Body Mass Index; Obesity; Varying-Coefficient; Variable Selection.

JEL classification: C13, C14, I15

The authors gratefully thank Susumu Imai from University of Technology Sydney for his constructive comments and suggestions.

- Jiti Gao, Department of Econometrics and Business Statistics, Monash University, VIC, Australia. Email: Jiti.Gao@monash.edu. Jiti Gao acknowledges the Australian Research Council Discovery Grants Program support under Grant numbers: DP130104229 and DP150101012.
- Bin Peng, Economics Discipline Group, University of Technology Sydney, NSW, Australia. Email: Bin.Peng@uts.edu.au.
- Zhao Ren, Department of Statistics, University of Pittsburgh, USA. Email: zren@pitt.edu.
- Xiaohui Zhang, School of Management and Governance, Murdoch University, Perth, Australia. Email: Xiaohui.Zhang@murdoch.edu.au.

1 Introduction

As a widely used measurement for body fat, body mass index (BMI) has been attracting significant attention from numerous researchers in multiple disciplines. The interest in measuring body fat came with increasing obesity in the last three decades especially in developed countries. According to WHO estimates, the worldwide prevalence of obesity more than doubled between 1980 and 2014. Overweight is a major risk factor for noncommunicable diseases including cardiovascular diseases such as hypertension and coronary disease, type II diabetes, gallbladder disease, musculoskeletal disorders especially osteoarthritis, and some cancers such as endometrial, breast and colon (Fontaine et al., 2003; WHO, 2015). The prevalence of these chronic diseases places a huge burden on health care and welfare systems (Mukherji et al., n.d.; Hu and Sasaki, 2015). It is thus crucial to identify and quantify the correlations between potential explanatory factors and BMI. Empirical studies, which try to link particular lifestyle behaviours and other risk factors to BMI, will inform and guide policy makers to provide efficient incentives and interventions to reduce population BMI. Numerous studies have been seen in the last two decades and a large number of factors have been proposed as important drivers of increasing BMI (for references see Cawley (2011)). While there is an impressive amount of evidence on the individual importance of determinants, there is little guidance for policy makers where to focus cost-containment efforts. Because very few, if not none, of these empirical studies can tell which of the many proposed drivers are important to BMI and which are not.

Another challenge in providing valid and efficient guidance to policy makers is that impacts of determinants on BMI vary across demographic groups. It is well documented in the literature. For example, a range of studies found impacts of socio-economic factors, lifestyles and marital status on BMI were different across age, gender and ethnic groups (Colditz et al., 1991; Sobal et al., 1992; Lipowicz et al., 2002; Zhang and Wang, 2004; Yu, 2012). However, almost all of the previous studies allowed for varying impact of one particular determinant rather than all determinants. The choice of determinants having varying impacts is arbitrary and lack of statistical support. The main objective of our study is to identify the most important drivers of BMI, and to quantify their varying impacts on BMI across demographic groups. To fulfill the objective, we propose a variable selection method for the so-called categorical varying-coefficient model discussed by Li et al. (2013).

Varying-coefficient models have attracted considerable attention and gained popularity in the past two decades from both theoretical and practical aspects (e.g. Hastie and Tibshirani, 1993; Fan and Zhang, 1999; Wang and Xia, 2009; Li and Racine, 2010; Gao and Phillips, 2013; Li et al., 2013; and so forth). This type of models has been widely used due to its good interpretability and flexibility. It generalizes the linear regression model, which keeps the rigid control through a parametric specification, and provides an automatic and flexible approach to the linear parameters through a nonparametric specification. Variable selection for varying-coefficient models has received increasing attention (Wang et al., 2008; Wang and Xia, 2009). To the best of our knowledge, almost all of these existing variable selection methods for varying-coefficient models are specifically for the setting that only continuous predictors or indexes enter the nonparametric specification of linear parameters. However it is very common in applied settings that categorical variables influence the regressors' impacts on dependent variable, such as our study in this paper.

Li et al. (2013) proposed a kernel based consistent estimation method for the varying-coefficient model with only categorical variables affecting the coefficients, and it has been widely applied (see, for example, Delgado (2013) and Hendricks and Smith (2015)). In Li et al. (2013), the consistency and asymptotic normality of the parameter function have been fully investigated and the choices of optimal bandwidths have also been studied in details. Based on their work, the confidence intervals can be obtained, so one can tell if the estimates on the corresponding parameters of interest are significant different from zero or not. However, insignificance does not imply one should completely remove the corresponding regressor from the system. One well known example is R^2, which is always employed to measure the goodness of fit of a model. Loosely speaking, R^2 shows a monotone increase with the number of variables included no matter if these variables have significant estimated coefficients. Although adjusted R^2 has been provided to solve variable selection problem for the parametric linear model, how to deal with this difficulty for semiparametric categorical varying-coefficient model remains unknown. As discussed in Wang and Xia (2009), including spurious regressors can degrade the estimation efficiency substantially. In order to address this issue, we implement a variable selection procedure for the categorical varying-coefficient model proposed in Li et al. (2013). The method is demonstrated by using both simulated and real data examples.

The methodological and theoretical contribution of the current paper is threefold. First, we propose a novel data-driven and practical optimization procedure to solve the variable selection problem for the categorical varying-coefficient model. Second, theoretical result is established that true model can be successfully detected, which cannot be obtained using the method proposed in Li et al. (2013). Last but not least, we derive that besides variable selection result, our estimator also achieves asymptotical normality on the true (oracle) model.

The rest of the paper is organized as follows. In Section 2, we review the models of Li et al. (2013) at first, and then introduce a variable selection method and its asymptotic results. In Section 3, we conduct a Monte Carlo study to investigate the finite sample property of the method. In Section 4, by using the 2013 National Health Interview Survey data in the U.S., we apply our method to identify determinants of BMI and to quantify their varying effects on BMI across demographic groups, including gender, age and ethnicity. Section 5 gives some conclusions and then discusses some issues related to the proposed model selection. An estimation procedure is given in Appendix A. Appendices B-D provide the summary statistics of the real data under study and then the estimation results. The mathematical proofs of the main results are given in Appendix E.

2 Econometric Framework

In this section, we review the model discussed in Li et al. (2013). Based on some fundamental settings and results of Li et al. (2013), we then introduce our variable selection procedure with corresponding theoretical properties.

2.1 Brief Review

The model of Li et al. (2013) is specified as follows.

$$Y_i = X_i'\beta_0(Z_i) + \varepsilon_i, \quad i = 1, \ldots, N, \tag{2.1}$$

where $Z_i = (\bar{Z}_i', \tilde{Z}_i')'$ is an r-dimensional vector of discrete covariates with a support $\mathcal{D} = \bar{\mathcal{D}} \times \tilde{\mathcal{D}}$, $\bar{Z}_i = (Z_{i,1}, \ldots, Z_{i,\bar{r}})'$, $\tilde{Z}_i = (Z_{i,\bar{r}+1}, \ldots, Z_{i,r})'$ and $1 \leq \bar{r} \leq r$. Moreover, $\{\tilde{Z}_i, 1 \leq i \leq N\}$ is independent of all other variables and has no impact on $\beta_0(\cdot)$, which implies that $\tilde{Z}_i$ has no impact on Y_i at all. Therein, $\bar{Z}_i$ and $\tilde{Z}_i$ are referred to as relevant and irrelevant covariates respectively. When $\bar{r} = r$, there is no irrelevant covariate existing in the system, i.e. $\bar{Z}_i = Z_i$. To distinguish X_i from Z_i, they are respectively referred to as regressors and covariates hereafter. Based on the above description, the true model reduces to

$$Y_i = X_i'\beta_0(\bar{Z}_i) + \varepsilon_i, \quad i = 1, \ldots, N, \tag{2.2}$$

where ε_i is a random error term; $X_i = (X_{i,1}, \ldots, X_{i,p})'$ is a p-dimensional vector of regressors; $\beta_0(z) = (\beta_{01}(z), \ldots, \beta_{0p}(z))'$ is a p-dimensional unknown coefficient function; and no information is known in advance to distinguish $\bar{Z}_i$ and $\tilde{Z}_i$. Moreover, both p and r are assumed to be fixed for the above model.

In order to retrieve the information hidden in Z_i (or remove some spurious information from Z_i), Li et al. (2013) proposed to use a kernel function to carry on regression. More specifically, the kernel function of Aitchison and Aitken (1976) for an unordered covariate is adopted.

$$l(Z_{i,s}, z_s, \theta_s) = \begin{cases} 1, & \text{if } Z_{i,s} = z_s \\ \theta_s, & \text{otherwise} \end{cases}, \tag{2.3}$$

where the range of θ_s is $[0, 1]$ for $s = 1, \ldots, r$. It can be seen that $\theta_s = 0$ leads to an indicator function and $\theta_s = 1$ gives a uniform weight function. Then (2.3) allows us to construct a product kernel function as follows:

$$L(Z_i, z, \Theta) = \prod_{s=1}^{r} l(Z_{i,s}, z_s, \theta_s) = \prod_{s=1}^{r} \theta_s^{1(Z_{i,s} \neq z_s)}, \tag{2.4}$$

where $\Theta = (\theta_1, \ldots, \theta_r)'$. Therefore, for $\forall z \in \mathcal{D}$, the kernel based OLS estimator is denoted as

$$\hat{\beta}(z) = \left(\sum_{j=1}^{N} X_j X_j L(Z_j, z, \hat{\Theta}) \right)^{-1} \sum_{j=1}^{N} X_j Y_j L(Z_j, z, \hat{\Theta}),$$

where the optimal bandwidth $\hat{\Theta}$ is obtained by minimizing the cross-validation criterion function

$$CV(\Theta) = \frac{1}{N} \sum_{i=1}^{N} \left(Y_i - X_i' \hat{\beta}_{-i} \right)^2 \tag{2.5}$$

and the leave-one-out OLS estimator $\hat{\beta}_{-i}$ is defined as

$$\hat{\beta}_{-i} = \left(\sum_{j=1, j \neq i}^{N} X_j X_j L(Z_j, Z_i, \Theta) \right)^{-1} \sum_{j=1, j \neq i}^{N} X_j Y_j L(Z_j, Z_i, \Theta).$$

Li et al. (2013) further establish two important results on relevant covariates: the bandwidths of relevant covariates converge to zero at least with the rate $O_P\left(\frac{1}{\sqrt{N}}\right)$; when there is no irrelevant covariate in the system, the bandwidths of relevant covariates converge to zero at $O_P\left(\frac{1}{N}\right)$. Since this paper is a further study of model (2.2) and we need to employ the optimal bandwidth obtained from (2.5) in the following analysis, we, for the convenience of readers, introduce some notation, and then state the assumptions and these two results of Li et al. (2013) as a lemma.

For an r-dimensional vector $z = (z_1, \ldots, z_r)' \in \mathcal{D}$, we partition z as $z = (\bar{z}', \tilde{z}')'$ conformably with Z_i, where $\bar{z} = (z_1, \ldots, z_{\bar{r}})'$ and $\tilde{z} = (z_{\bar{r}+1}, \ldots, z_r)'$. Correspondingly, we partition Θ as $\Theta = (\bar{\Theta}', \tilde{\Theta}')'$, where $\bar{\Theta} = (\theta_1, \ldots, \theta_{\bar{r}})'$ and $\tilde{\Theta} = (\theta_{\bar{r}+1}, \ldots, \theta_r)'$.

Since the following two assumptions are identical to those in Li et al. (2013), so all the relevant discussions are omitted.

Assumption 1:

1. Let $\{X_i, Z_i, Y_i\}_{i=1}^N$ be independent and identically distributed. Let $\max_{\bar{z} \in \bar{\mathcal{D}}} \|\beta_0(\bar{z})\| < \infty$.

2. $E[Y_i^2 | X_i = x, \bar{Z}_i = \bar{z}]$ is bounded on $(x, \bar{z}) \in \mathbb{R}^p \times \bar{\mathcal{D}}$.

3. Defining $\sigma_\varepsilon^2(x, \bar{z}) = E[\varepsilon_i^2 | X_i = x, \bar{Z}_i = \bar{z}]$ and $\sigma_\varepsilon^2(\bar{z}) = E[\sigma_\varepsilon^2(X_i, \bar{z}) | \bar{Z}_i = \bar{z}]$, then $E[\sigma_\varepsilon^2(X_i, \bar{z}) X_i X_i' | \bar{Z}_i = \bar{z}]$ is positive definite for all $\bar{z} \in \bar{\mathcal{D}}$.

4. For $s = 1, \ldots, r$, the sth component of $z = (z_1, \ldots, z_r)'$ (i.e. z_s) takes c_s different values in $\{0, 1, \ldots, c_s - 1\}$. Moreover, $2 \leq \min_{1 \leq s \leq r} c_s \leq \max_{1 \leq s \leq r} c_s < \infty$.

Assumption 2:

1. **Relevant Covariate Case: i.e.** $\bar{r} = r$
   Define $L_{ij,\Theta} = L(Z_i, Z_j, \Theta)$, $\eta_\beta(Z_j) = (E[X_iX_i'L_{ij,\Theta}|Z_j])^{-1}E[X_iX_i'\beta(Z_i)L_{ij,\Theta}|Z_j]$ and $m(Z_i) = E[X_iX_i'|Z_i]$. Then the only values of $\Theta = (\theta_1, \ldots, \theta_r)'$ that make

$$\sum_{z \in \mathcal{D}} \Pr(z)[\eta_\beta(z) - \beta_0(z)]'m(z)[\eta_\beta(z) - \beta_0(z)] = 0 \tag{2.6}$$

   are $\Theta = 0_{r\times 1}$.

2. **Irrelevant Covariate Case: i.e.** $\bar{r} < r$
   Define $\bar{L}_{ij,\bar{\Theta}}$, $\bar{\eta}_\beta(\bar{Z}_j) = (E[X_iX_i'\bar{L}_{ij,\bar{\Theta}}|\bar{Z}_j])^{-1}E[X_iX_i'\beta(\bar{Z}_i)\bar{L}_{ij,\bar{\Theta}}|\bar{Z}_j]$ and $\bar{m}(\bar{Z}_i) = E[X_iX_i'|\bar{Z}_i]$. Then the only values of $\Theta = (\theta_1, \ldots, \theta_{\bar{r}})'$ that make

$$\sum_{\bar{z} \in \bar{\mathcal{D}}} \Pr(\bar{z})[\bar{\eta}_\beta(\bar{z}) - \beta_0(\bar{z})]'m(\bar{z})[\bar{\eta}_\beta(\bar{z}) - \beta_0(\bar{z})] = 0 \tag{2.7}$$

   are $\bar{\Theta} = 0_{\bar{r}\times 1}$. $\theta_s \in [0, 1]$ for $s = \bar{r} + 1, \ldots, r$.

For detailed discussion on Assumptions 1 and 2, please see Li et al. (2013).

Lemma 2.1. *Let* $\hat{\Theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_r)' = \operatorname{argmin}_{\Theta \in [0,1]^p} CV(\Theta)$.

1. *Under Assumptions 1 and 2.1,* $\hat{\theta}_s = O_P\left(\frac{1}{N}\right)$ *for* $s = 1, \ldots, r$.

2. *Under Assumptions 1 and 2.2,* $\hat{\theta}_s = O_P\left(\frac{1}{\sqrt{N}}\right)$ *for* $s = 1, \ldots, \bar{r}$, *and* $\lim_{N\to\infty} \Pr(\hat{\theta}_{\bar{r}+1} = 1, \ldots, \hat{\theta}_r = 1) \geq \alpha$ *for some* $\alpha \in (0, 1)$.

Lemma 2.1 summaries Theorems 1 and 3 of Li et al. (2013) and provides all the information that requires for the covariates Z_i. Based on this result, consistent estimation on $\beta_0(\cdot)$ can then be established. Since those estimation details are not directly related to this paper, we will not discuss them here.

Notice that if for certain covariate we obtain $\hat{\theta}_s = 1$, then we can safely remove the corresponding covariate from the system. Although one cannot always achieve $\hat{\theta}_s = 1$ for all irrelevant covariates simultaneously, Lemma 2.1 implies that there is always certain positive probability that we can recognize all irrelevant covariates. In view of this, the above result provides a variable selection procedure for the elements of covariates Z_i.

With these optimal bandwidths, we are able to propose a variable selection procedure for the elements of regressors X_i, which is the main focus of this paper. Hereafter, with slight abuse of notation, we assume that we have removed all detected (those $Z_{i,s}$ with $\hat{\theta}_s = 1$) irrelevant covariates according to Lemma 2.1 and the remaining covariates of the ith observation is still named $Z_i = (\bar{Z}_i', \tilde{Z}_i')'$ as before. However, it is clear that there is positive probability such that no $\tilde{Z}_i$ exists. The purpose of this step is to reduce the total number of distinct realizations of z from our samples $\{Z_1, \ldots, Z_N\}$, which is denoted as m in the next subsection.

2.2 Variable Selection on X_i

In this section we propose a variable selection procedure for the model (2.2). The main objective is to select the variables of X_i with nonzero coefficients when p and r are fixed. To this end, we assume that there exists an unknown set $\mathcal{A} \subseteq \{1, \ldots, p\}$ satisfying that $E|\beta_{0j}(\bar{Z}_i)|^2 = 0$ if and only if $j \in \mathcal{A}$, where $\beta_{0j}(\bar{Z}_i)$ denotes the jth element of $\beta_0(\bar{Z}_i)$. For simplicity of the notation, in the rest of this paper we assume that in the true model, $\mathcal{A} = \{p^* + 1, \ldots, p\}$ for some positive integer $p^* \leq p$. In other words, only the first p^* variables in X_i have nonzero coefficients and our goal is to find this unknown $\mathcal{A}$.

Let m denote the number of different realizations of z by observing $\{Z_1, \ldots, Z_N\}$. It is obvious that m converges to the cardinality of $\mathcal{D}$ in probability with non-degenerate probability imposed on i.i.d. Z_i as N diverges to ∞.

- **Remark:** Obviously, the cardinality of $\mathcal{D}$ is no more than $c_1 \times \cdots \times c_r$, where, for $s = 1, \ldots, r$, c_s is the number of different values for categorical variable $Z_{i,s}$ and denoted in Assumption 1. m always achieves the cardinality of $\mathcal{D}$ with probability 1 as long as N is sufficiently large. In this sense, it suits our empirical study well since the number of individuals is much larger than that of the groups generated according to age, gender and race. See further details in Section 4.

Since m is finite and observable, our parameters of interest can be characterized by the following $m \times p$ matrix B with the underlying true coefficient function B_0. For the sake of presentation, denote

$$\begin{aligned}
\underset{m\times p}{B} &= (\beta_1, \ldots, \beta_m)' = (b_1, \ldots, b_p), \\
\underset{p\times 1}{\beta_j} &= (\beta_{j,1}, \ldots, \beta_{j,p})' \text{ for } j = 1, \ldots, m, \\
\underset{m\times 1}{b_s} &= (\beta_{1,s}, \ldots, \beta_{m,s})' \text{ for } s = 1, \ldots, p, \\
\underset{m\times p}{B_0} &= (\beta_0(z^1), \ldots, \beta_0(z^m))' = (b_{01}, \ldots, b_{0p^*}, 0, \ldots, 0), \\
\underset{m\times 1}{b_{0s}} &= (\beta_{0s}(z^1), \ldots, \beta_{0s}(z^m))' \text{ for } s = 1, \ldots, p^*, \qquad (2.8)
\end{aligned}$$

where z^j for $j = 1, \ldots, m$ denotes the jth realization of $z \in \mathcal{D}$.

Notice that the last $p - p^*$ columns of B_0 are zeros, which implies a group structure in B_0. In other words, entries in each column of B_0 form a group. Then the selection on variables becomes identifying those 0 columns in the matrix B_0. Following the spirit of Yuan and Lin (2006), we consider the following regularized least squares estimator:

$$\hat{B} = (\hat{\beta}_{\gamma,1}, \ldots, \hat{\beta}_{\gamma,m})' = (\hat{b}_{\gamma,1}, \ldots, \hat{b}_{\gamma,p}) = \operatorname*{argmin}_{B \in \mathbb{R}^{m \times p}} Q_\gamma(B) \tag{2.9}$$

and

$$Q_\gamma(B) = \sum_{j=1}^{m} \sum_{i=1}^{N} (Y_i - X_i'\beta_j)^2 \, L(Z_i, z^j, \hat{\Theta}) + \sum_{s=1}^{p} \gamma_s \|b_s\|, \tag{2.10}$$

where $\hat{\Theta}$ is obtained from Lemma 2.1; b_s is the sth column of B for $s = 1, \ldots, p$ and denoted in (2.8); the term $\sum_{s=1}^{p} \gamma_s \|b_s\|$ is the group-wise regularizer and is defined as the weighted sum of the ℓ_2 norms of all the column vectors in B with the weight $\gamma = (\gamma_1, \ldots, \gamma_p)'$ controlling the regularizer.[1]

In order to derive our theorems, we further impose the following conditions.

Assumption 3:

1. For a random variable $\bar{Z}_i \in \bar{\mathcal{D}}$ and $\beta_0(\bar{Z}_i) = (\beta_{01}(\bar{Z}_i), \ldots, \beta_{0p}(\bar{Z}_i))'$, suppose there exists a positive integer $p^* \leq p$ such that $0 < E|\beta_{0j}(\bar{Z}_i)|^2 < \infty$ for $j = 1, \ldots, p^*$ and $E|\beta_{0j}(\bar{Z}_i)|^2 = 0$ for $j = p^* + 1, \ldots, p$.

2. For $\forall \bar{z} \in \bar{\mathcal{D}}$, $0 < \alpha_1 \leq \rho_{min} \leq \rho_{max} \leq \alpha_2 < \infty$, where ρ_{min} and ρ_{max} denote the minimum and maximum eigenvalues of $E[X_i X_i' | \bar{z}]$ respectively, and α_1, α_2 are two universal positive constants. For $\forall z \in \mathcal{D}$, $\Pr(z) = \Pr(Z_i = z) > \alpha_3 > 0$ with some universal constant α_3.

Assumption 3.2 ensures both minimum and maximum eigenvalues of $E[X_i X_i' | \bar{z}]$ are bounded uniformly in N.

We are now ready to establish the main results in Theorems 2.1-2.3 below; their proofs are provided in Appendix E.

Theorem 2.1. *Let Assumptions 1-3 hold.*

1. *Let* $\gamma^* = (\gamma_1, \ldots, \gamma_{p^*})'$ *and* $\frac{\|\gamma*\|}{\sqrt{N}} \to \omega_1$*, where* ω_1 *is a constant satisfying* $0 \leq \omega_1 < \infty$*. Then*

$$\left\| \hat{\beta}_{\gamma,j} - \beta_0(\bar{z}^j) \right\| = O_P\left(N^{-1/2}\right)$$

for $j = 1, \ldots, m$*, where* $\bar{z}^j = (z_1^j, \ldots, z_{\bar{r}}^j)'$.

2. *Let* $\frac{1}{\sqrt{N}} \min_{s \in \{p^*+1, \ldots, p\}} \gamma_s \geq \omega_2$*, where* ω_2 *is a sufficiently large constant. Then*

$$\Pr(\|\hat{b}_{\gamma,j}\| = 0) \to 1$$

[1] In the literature of group LASSO analysis, one normally allows both p and r to diverge to infinity (e.g. Lounici et al. (2011)). However, to the best knowledge of authors, how to achieve the optimal bandwidths for model (2.2) remains unknown for high dimensional cases. Given that this note is a complementary study of Li et al. (2013) and finite p and r suit our case study better, we will not pursue the situations where both p and r diverge to infinity.

for $j = p^* + 1, \ldots, p$.

The first result of Theorem 2.1 says if the regularizer weight is not too large, we always have optimal $\sqrt{N}$ consistency for our estimator. The second result implies that when the regularizer weight is at level $\sqrt{N}$, we can successfully get rid of those unimportant coefficients in our estimator and select a sub-model of the true model. A natural and simple choice of γ, which satisfies assumptions of both results, is that all the elements of γ are at level $\sqrt{N}$. However, with such choice of γ, Theorem 2.1 does not imply the asymptotic normality property of our estimator for the selected model, such as that obtained in Li et al. (2013) for the oracle estimator (i.e. the unregularized estimator obtained under the true model) defined below,

$$\hat{\beta}_{ora}(\bar{z}^j) = \left(\sum_{i=1}^{N} X_{iU} X_{iU}' L(Z_i, z^j, \hat{\Theta}) \right)^{-1} \sum_{i=1}^{N} X_{iU} Y_i L(Z_i, z^j, \hat{\Theta}), \tag{2.11}$$

where $j = 1, \ldots, m$ and $X_{iU} = (X_{i,1}, \ldots, X_{i,p^*})'$. Note that the asymptotic properties of the oracle estimator have been fully examined in Li et al. (2013), wherein all the references can be found.[2]

In fact, with a more careful data-driven choice of γ, we can further achieve the asymptotic normality whenever there is no irrelevant covariate with the help of following oracle property for our estimator (2.9).

Theorem 2.2. *Under the conditions of Theorem 2.1,*

$$\left\| \hat{\beta}_{\gamma,jU} - \hat{\beta}_{ora}(\bar{z}^j) \right\| = O_P\left(\frac{\|\gamma^*\|}{N} \right)$$

for $j = 1, \ldots, m$, *where* $\hat{\beta}_{\gamma,jU} = (\hat{\beta}_{\gamma,j1}, \ldots, \hat{\beta}_{\gamma,jp^*})'$; $\hat{\beta}_{\gamma,js}$ *denotes the sth element of* $\hat{\beta}_{\gamma,j}$ *for* $j = 1, \ldots, m$ *and* $s = 1, \ldots, p^*$; *and* γ^* *is denoted in Theorem 2.1.*

In order to achieve an asymptotic normality for the selected model, the rate of convergence of $\hat{\beta}_{\gamma,jU}$ to $\hat{\beta}_{ora}(\bar{z}^j)$ should be much faster than $\frac{1}{\sqrt{N}}$. The oracle property in Theorem 2.2 implies such a result as long as $\|\gamma^*\|$ is much smaller than $\sqrt{N}$. Therefore the simple choice of $\sqrt{N}$ level for γ suggested above is not sufficient to achieve an asymptotic normality. In the following, we propose a data-driven choice of γ, which can yield a even faster rate $O_P\left(\frac{1}{N}\right)$ of convergence to the oracle and thus achieve the desired asymptotic normality property. From now on, we assume that whenever the true coefficient is nonzero, that is $b_{0s} \neq 0$ for $s = 1, \ldots, p^*$, its ℓ_2 norm is larger than some universal constant $\|b_{0s}\| > \alpha_0 > 0$. This assumption is natural in the current fixed dimension setting.

[2]Notice that the word "oracle" refers to those estimators provided in Li et al. (2013) by assuming we know the true set $\mathcal{A}$. Here we completely ignore the inefficiency brought in the model by the irrelevant covariates $\tilde{Z}_i$. The asymptotically efficient estimator is obtained when we know both the set $\mathcal{A}$ and the irrelevant covariates. However, this can only be done at certain probability based on Lemma 2.1.

Similar to Wang and Leng (2007) and Wang and Xia (2009), we pick our data-driven regularizer weight as follows,

$$\gamma = \tilde{\gamma}\left(\|\tilde{b}_1\|^{-1}, \ldots, \|\tilde{b}_p\|^{-1}\right)', \tag{2.12}$$

where $\tilde{\gamma}$ is a scalar, $\tilde{b}_s$ is the sth column of the unregularized estimator $\tilde{B}$, and $\tilde{B}$ is obtained from (2.10) by simply choosing $\gamma_1 = \cdots = \gamma_p = 0$. In connection with Assumption 3.1, the first result of Theorem 2.1 and our assumption $\|b_{0s}\| > \alpha_0 > 0$ for $s = 1, \ldots, p^*$, it is easy to verify that $\|\tilde{b}_s\|^{-1} = O_P(1)$ for $s = 1, \ldots, p^*$ and $\|\tilde{b}_s\| = O_P\left(\frac{1}{\sqrt{N}}\right)$ for $s = p^* + 1, \ldots, p$. Then the intuition of choosing γ in (2.12) is straightforward. Indeed, the unregularized estimator $\tilde{B}$ provides us desired $\sqrt{N}$ consistent estimation. We thus can take advantage of this information and know if each column of B_0 is likely to be zero or not. In other words, smaller $\|\tilde{b}_j\|$ implies that the jth column is more likely to be zero and hence suggests a larger regularizer on $\|b_j\|$. Given the form of γ in (2.12), the selection on the vector γ becomes the selection on the scalar $\tilde{\gamma}$. Note that the properties of $\|\tilde{b}_j\|^{-1}$ for $j = 1, \ldots, p$ imply that a large enough constant $\tilde{\gamma}$ would satisfy all the technical conditions on γ needed for the above theorems with $\left\|\hat{\beta}_{\gamma,jU} - \hat{\beta}_{ora}(\bar{z}^j)\right\| = O_P\left(\frac{1}{N}\right)$. More specifically, we select the constant $\tilde{\gamma}$ by the following modified BIC-type (MBIC) criterion.

$$MBIC_{\tilde{\gamma}} = \ln RSS_{\tilde{\gamma}} + df_{\tilde{\gamma}} \cdot \frac{\ln N}{N},$$

where $df_{\tilde{\gamma}}$ is simply the number of nonzero coefficients identified by $\hat{B}_{\tilde{\gamma}}$ and $RSS_{\tilde{\gamma}}$ is defined as

$$RSS_{\tilde{\gamma}} = \frac{1}{N}\sum_{j=1}^{m}\sum_{i=1}^{N}\left(Y_i - X_i'\hat{\beta}_{\tilde{\gamma},j}\right)^2 L(Z_i, z^j, \hat{\Theta}).$$

The weight parameter is obtained by

$$\hat{\tilde{\gamma}} = \underset{\tilde{\gamma}}{\operatorname{argmin}}\, MBIC_{\tilde{\gamma}}. \tag{2.13}$$

Recall we assume the true set of nonzero coefficients is denoted by $\mathcal{A}^c = \{1, \ldots, p^*\}$. Let's set $S_{\hat{\tilde{\gamma}}} = \{j : \|\hat{b}_{\hat{\tilde{\gamma}},j}\| > 0, 1 \le j \le p\}$ as the variables identified by the regularized estimator $\hat{B}_{\hat{\tilde{\gamma}}}$ using the weight parameter $\hat{\tilde{\gamma}}$ picked in (2.13). Then the next result follows.

Theorem 2.3. *Under conditions of Theorem 2.1, the weight parameter selected by the modified BIC-type criterion (2.13) can:*

1. *Identify the true model consistently, i.e.* $\Pr(S_{\hat{\tilde{\gamma}}} = \mathcal{A}^c) \to 1$ *as* $N \to \infty$*;*

2. *For relevant covariate case defined in Assumption 2, achieve the asymptotic normality, i.e.*

$$\sqrt{N}(\hat{\beta}_{\hat{\tilde{\gamma}},jU} - \beta_0(z^j)) \to_D N(0, \Sigma(z^j)) \tag{2.14}$$

for $j = 1, \ldots, m$, *where* $\Sigma(z^j) = A^{-1}(z^j)\Omega(z^j)A^{-1}(z^j)$, $A(z^j) = E[X_i X_i | z^j] \Pr(z^j)$ *and* $\Omega(z^j) = E[\varepsilon_i^2 X_i X_i | z^j] \Pr(z^j)$.

3. *For irrelevant covariate case defined in Assumption 2,*

$$\hat{\beta}_{\hat{\bar{\gamma}},jU} - \beta_0(\bar{z}^j) = O_P\left(\frac{1}{\sqrt{N}}\right) \tag{2.15}$$

for $j = 1, \ldots, m$,

When there is no irrelevant covariate (i.e. $r = \bar{r}$ and $Z_i = \bar{Z}_i$), the asymptotic normality result of (2.14) is based on the limiting distribution of $\sqrt{N}(\hat{\beta}_{ora}(\bar{z}^j) - \beta_0(\bar{z}^j))$, which is established by applying Theorem 2 of Li et al. (2013) on the oracle model. However, when there are irrelevant covariates (i.e. $r > \bar{r}$), the asymptotic distribution of $\sqrt{N}(\hat{\beta}_{ora}(\bar{z}^j) - \beta_0(\bar{z}^j))$ remains unknown even for the oracle estimator and hence we only obtain $\sqrt{N}$ consistency in (2.15).

In summary, in this section, we introduce a regularized approach for the categorical varying-coefficient model and obtain superior statistical properties for the variable selection method. In particular, for the categorical model, the coefficients demonstrate a natural group structure. To take advantage of the structure, we apply a group-wise regularizer to improve our estimation accuracy and select the true model successfully. Moreover, we apply a data-driven method to select the weight parameter using the modified BIC-type criterion to further boost the performance and achieve an asymptotic normality result wherever no irrelevant covariate exists.

3 Monte Carlo Evidence

In this section, we conduct a Monte Carlo study to investigate the finite sample properties of our method. The detailed estimation procedure is provided in Appendix A.

The data generating process (DGP) is as follows.

$$Y_i = X_i'\beta_0(Z_i) + \varepsilon_i \quad \text{and} \quad X_i = H_i + V_i.$$

Let $p = 5$, $p^* = 2$ and $r = 2$. Let $Z_i = (Z_{i,1}, \cdots, Z_{i,r})'$, where for $\forall j = 1, \ldots, r$ $Z_{i,j}$ is i.i.d. over i; and $Z_{i,j}$ is chosen from $\{0, 1, 2, 3\}$ with the same probability all the time, i.e. $\Pr(Z_{i,j} = 0) = \Pr(Z_{i,j} = 1) = \Pr(Z_{i,j} = 2) = \Pr(Z_{i,j} = 3) = 0.25$. V_i is i.i.d. over i and follows $N(Z_{i,1}/2 \cdot i_p, \sqrt{Z_{i,1} + 1} \cdot I_p)$. H_i is i.i.d. over i and follows $N(i_p, I_p)$. ε_i is i.i.d. over i and follows $N(0, 1)$. Let $\beta_{0j}(Z_i)$ denote the jth element of the coefficient function $\beta_0(Z_i)$ for $j = 1, \ldots, p$.

We consider both relevant and irrelevant covariate cases respectively as follows.

- Relevant Covariate (i.e. $\bar{r} = r = 2$): For $\forall j = 1, 2$, $\beta_{0j}(Z_i) = j/2 \cdot \sum_{k=1}^{r} Z_{i,k} + 5$; for $j > 2$, $\beta_{0j} = 0$.

- Irrelevant Covariate (i.e. $r = 2$ and $\bar{r} = 1$): For $\forall j = 1, 2$, $\beta_{0j}(Z_i) = j/2 \cdot Z_{i,1} + 5$; for $j > 2$, $\beta_{0j} = 0$.

For each generated data set, we estimate $\hat{B}$ by (2.9). Then we record the bias and squared error for every element of $\hat{B}$ in each replication. After 1000 replications, we calculate the averaged bias and squared errors for every element of $\hat{B}$. We refer to the averaged bias and MSE matrices as BB and MB. Finally, we sum up the absolute biases and mean squared errors as follows.

$$\text{Bias} = \frac{1}{p} \sum_{s=1}^{p} \sum_{j=1}^{m} \Pr(z^j) \left| \text{BB}_{js} \right| \quad \text{MSE} = \frac{1}{p} \sum_{s=1}^{p} \sum_{j=1}^{m} \Pr(z^j) \text{MB}_{js}, \tag{3.1}$$

where BB_{js} and MB_{js} represent the (j, s)th elements for matrices BB and MB respectively, and z^j is denoted in (2.10). As discussed under (2.12), in order to ensure the requirements on the weight parameter hold, we simply use $\tilde{\gamma} = 50$ throughout the simulation. Setting $\tilde{\gamma}$ as a fixed number is purely for reducing the computational burden for the Monte Carlo study. In the empirical study, we choose the opimal $\tilde{\gamma}$ by using (2.13).

For comparison, we also apply the post-selection estimator (i.e. applying the unregularized estimator to the model excluding irrelevant regressors selected by (2.9)), the unregularized estimator (i.e. applying the unregularized estimator to the model including all regressors), and the oracle estimator. For each of these three estimators, we calculate Bias and MSE in the same fashion as (3.1).

Given the above set-up, we expect to obtain 3 columns of 0 by using (2.9) for each replication. Thus, we report the probability of the true model being identified in 1000 replications. As Table 1 shows, for both relevant and irrelevant covariate cases, Bias and MSE decrease as the sample size increases. The same argument is applicable to the accuracy rate of the true model being identified. It is not surprising to see that the oracle estimator has the best performance. Although (2.9) always has the largest MSE, the post-selection estimator provides much more accurate estimations.

4 An Application to BMI

In this section, we apply our variable selection method for the categorical varying-coefficient model to select the determinants from a wide range of potential factors and to quantify the varying impacts of the selected determinants by demographic characteristics.

4.1 Data

Data used in this empirical study are from the 2013 National Health Interview Survey (NHIS) in the United States. The NHIS is conducted annually through face-to-face interviews. Our

Table 1: Estimates and accurate rates of the true model being identified

	N	Relevant Case Bias	Relevant Case MSE	Relevant Case Acc rate	Irrelevant Case Bias	Irrelevant Case MSE	Irrelevant Case Acc rate
While-selection	200	0.0152	0.1112	0.8750	0.0375	0.1761	0.8720
	400	0.0052	0.0192	0.9810	0.0040	0.0219	0.9410
	600	0.0030	0.0067	0.9980	0.0038	0.0242	0.9560
	800	0.0022	0.0052	0.9990	0.0028	0.0124	0.9770
Post-selection	200	0.0069	0.0139		0.0020	0.0027	
	400	0.0026	0.0044		0.0007	0.0012	
	600	0.0015	0.0027		0.0006	0.0007	
	800	0.0014	0.0020		0.0005	0.0006	
Unregularized	200	0.0195	0.0491		0.0034	0.0066	
	400	0.0075	0.0148		0.0015	0.0030	
	600	0.0049	0.0088		0.0010	0.0020	
	800	0.0035	0.0062		0.0009	0.0015	
Oracle	200	0.0053	0.0108		0.0016	0.0022	
	400	0.0024	0.0043		0.0005	0.0011	
	600	0.0015	0.0027		0.0004	0.0007	
	800	0.0014	0.0020		0.0004	0.0005	

"While-selection" refers to the results obtained by (3.1) using the estimates from (2.9).

analysis is focused on adults aged 18 and over. BMI is calculated based on self-reported height and weight. Given our preliminary interest is in overweight and obesity, to avoid disturbance from the other negative side, we exclude underweight individuals[3], i.e. those with BMI less than 18.5, from our analysis. Because the BMI scores are skewed towards higher values, we use natural logarithm transformed BMI in our analysis (Dardanoni et al., 2011). The summary statistics of the real data and then the estimation results are given in Appendices B-D below.

Through a systematic review of literature in overweight and obesity, we test 23 factors' correlations with BMI in this empirical study.[4] The first group of factors we test is lifestyles, including physical activity, alcohol consumption and smoking habits. Physical activity has been identified as a safe and efficient obesity intervention by many studies in public health and epidemiology (see, for example, Galani and Schneider (2007)). In this study we measure physical activity at three intensity levels, i.e. vigorous activity, light/moderate activity and strength activity, since previous studies found that impacts of physical activity at different intensity levels on BMI are different. We also use frequency of computer usage as a proxy for sedentary behaviours. Colditz et al. (1991) investigated the relation between alcohol intake and BMI, and they found alcohol intake changed the amount of other energy intake and then

[3]This accounts for a very small proportion, i.e. 1.8 per cent, of the whole sample.

[4]The number of factors tested is restricted by information available in the data set.

impacted on BMI. Cawley and Scholder (2013) provided evidence that the demand for cigarettes was derived from the demand for weight control. A similar negative impact of smoking on BMI was found in Colditz et al. (1991). In this analysis we take into account both status and consumption rate of drinking and smoking.

Secondly, there is a substantial body of studies that demonstrates an negative relationship between socio-economic positions and obesity risk. For example Cohen et al. (2013) reviewed 289 papers examining the relationship between educational status and obesity in multiple disciplines including public health, psychology, education, economics, sociology and demography, and found most studies identified a consistent relationship. El-Sayed et al. (2012) reviewed 27 studies conducting cross-sectional analysis on relations between occupational social class (OSC) and obesity in the UK, which suggested low OSC was associated with a higher risk of obesity. Working arrangements and working hours were two more factors proposed in the literature having impacts on BMI. For example, Au et al. (2013) found that compared to employed women those out of the labour force or unemployed were more likely to lose weight and that working longer hours was associated with higher weight gain. Income is an important indicator of socio-economic status, and proxy indicators, such as house ownership, are often used in the previous studies as well (El-Sayed et al., 2012). Previous studies also found that having health insurance was associated with higher body mass in the U.S. by isolating the effects of moral hazard where people with health insurance may change their behaviors towards weight control (see, for example, Kelly and Markowitz (2009)). Along this line health service utilization may be correlated with BMI in the same direction. In this study we consider two measures of medical service utilization, i.e. consulting any health professional in the last 12 months or not and health care expenditure.

In addition, we control some other variables proposed in the literature having correlations with BMI or obesity, including marital status, duration of US residence, geographic location and depression. Marital status has long been recognized as a factor affecting not only morbidity and mortality but also risk of obesity and increasing BMI (Sobal et al., 1992; Lipowicz et al., 2002). Strong evidence also suggests that recent immigrants to high income countries tend to have lower BMI than the host population on arrival (Oza-Frank and Cunningham, 2010), but this initial health advantage erodes over time (Antecol and Bedard, 2006). As for the relation between obesity and depression, Faith et al. (2011) reviewed 15 studies testing the pathway of 'depression to obesity' and found more than half of these studies report significant associations to support the hypothesis that obesity was a possible consequence of depression.

Previous studies showed that the relations between BMI and its determinants vary across age, gender and ethnicity groups. For example, Colditz et al. (1991) found that the associations between alcohol intake and BMI varied across genders. In particular alcohol intakes impacted on female BMI negatively, however for males there was no such association. Yu (2012) found education attainment had different impacts on BMI in different gender, age and race groups, i.e. compared with college graduates, less educated whites and younger black women were more likely to be obese, and the differentials were larger for women than men, and weak or not existent among black men and older black women. The analysis from Zhang and Wang (2004) also suggested the degree of socio-economic inequality in obesity varied considerably across gender, age, and ethnic groups. The marital role appeared to impact on obesity and BMI differently for men and women, i.e. married men were more likely to be obese than never married or previously married men but marital status was not significantly associate with women's obesity (Sobal et al., 1992). Lipowicz et al. (2002) found that differences in the BMI between married and never married individuals increased with age. Therefore in the model specification for this empirical study, demographic factors, including age, gender and ethnicity, enter the nonparametric functions of coefficients to accommodate the assumption that impacts of potential factors on BMI vary over demographic groups. Table 2 lists all 32 (i.e. $m = 32$) possible realizations of the covariates of gender, age and ethnicity. Definitions and summary statistics of all variables included in this study are presented in Appendix B because of space limitation.

4.2 Results

4.2.1 Variable Selection

As mentioned in Section 2, we implement (2.5) to estimate the optimal bandwidth at first. Results on bandwidth reported in Table 3 reveal that all three covariates are relevant, however their influences on the impacts of regressors on BMI are quite different. In particular, ethnicity and gender have relatively stronger influences than age because the smoothing parameters associated with *race* and *sex* are much smaller than that of *age*.

With these smoothing parameters in hand we apply our method to select the relevant and irrelevant regressors to BMI. The optimal weight parameter selected by the modified BIC-type criterion through (2.13) is $\hat{\hat{\gamma}} = 3.2$. Table 4 presents the result of variable selection through equation (2.9). 24 regressors, out of 48 in total, are identified as relevant, while the others are irrelevant to BMI.

In particular, our estimate suggests that exercise is correlated with BMI, however the level of intensity and frequency does matter. For example, compared to never doing vigorous (or strength) activity, doing such a level of exercise less than once per week has almost no effect on BMI, while doing it more than once per week starts to change BMI. In terms of light/moderate activity, however, people have to do it more than three times per week to see some effect on BMI. Results from our study may provide guidance for policy makers to adopt more efficient incentives to avoid overweight or obesity, i.e. encouraging people to do more intensive exercise or

Table 2: List of realizations of covariates in the data

	Sex		Age				Ethnicity			
Group Index	M	F	<25	[25, 45)	[45, 65)	≥65	W	B	A	O
1	x		x				x			
2	x		x					x		
3	x		x						x	
4	x		x							x
5	x			x			x			
6	x			x				x		
7	x			x					x	
8	x			x						x
9	x				x		x			
10	x				x			x		
11	x				x				x	
12	x				x					x
13	x					x	x			
14	x					x		x		
15	x					x			x	
16	x					x				x
17		x	x				x			
18		x	x					x		
19		x	x						x	
20		x	x							x
21		x		x			x			
22		x		x				x		
23		x		x					x	
24		x		x						x
25		x			x		x			
26		x			x			x		
27		x			x				x	
28		x			x					x
29		x				x	x			
30		x				x		x		
31		x				x			x	
32		x				x				x

M = Male, F = Female

W = White, B = Black, A = Asian, O = Other

Table 3: Bandwidth for covariates

Covariate	Bandwidth
sex	0.1158
age	0.1979
race	0.0703

Table 4: List of relevant and irrelevant variables to BMI

Relevant variable	**Irrelevant variable**
lifestyle factors	*lifestyle factors*
vig_l2	vig_l1
vig_l3	mod_l1
mod_l3	mod_l2
str_l2	str_l1
str_l3	smk_sd
smk_ed	cpuse_1
smk_f	cpuse_2
cigsday	*socio-economic factors*
alc1yr	occup3
alc_life	occup4
alc_c1	working
alc_c2	unemp
alc_c3	nowork
alc_c4	wrkhrs
socio-economic factors	houseown
educ1	notcov
occup1	hce_l2
occup2	hce_l3
lnincome	hce_l4
hp	hce_l5
other factors	hce_l6
us_born	*other factors*
us_m15	us_m5l15
rg_sth	citizenp
married	rg_ne
mental	rg_mw

to do moderate exercise more frequently rather than simply promoting exercise at any intensive level with any frequency.

Both the status of drinking and smoking and their consumption level are relevant to BMI. No impact from computer use can be seen. For socio-economic factors, education, income, and the two highest levels of OSC (*occup1* and *occup2*, compared to lowest OSC, i.e. *occup5*), and health professional visit in the last 12 months are identified as relevant regressors for BMI, but the two lower levels of OSC (*occup3* and *occup4*, compared to *occup5*), working arrangement, working hours, house ownership, health insurance coverage and medical care expenditure are irrelevant to BMI. Among the other factors, indicators on duration of living in the U.S. (i.e. born in the U.S. and living in the U.S. more than 15 years, compared to living in the U.S. less than 5 years), living in the south (compared to living in the west), marital status and mental health problems are robust factors for BMI, however living in the US more than 5 years but less than 15 years (compared to less than 5 years), citizenship, living in either the north east or the middle west (compared to living in the west) have no impact on BMI.

For comparison, we estimate the model using the unregularized estimator without variable selection. The full estimation results are reported in Appendix C 1-3 for the sake of space limitation. Table 5 lists all potential determinants of BMI considered in this study and the number of groups for whom the corresponding determinant has significant impact on BMI at 95% confidence level. Loosely speaking, a regressor can be identified as irrelevant regressor to BMI by unregularized estimator only when it has no significant impact on BMI for any demographic group. Two results emerge from Table 5. First, irrelevant regressors identified by unregularized estimator are also identified as irrelevant by regularized estimator. In particular, there are only three regressors, i.e. *cpuse_1*, *unemp* and *citizenp*, identified as irrelevant regressor by the unregularized estimator. For example, Figure 1 shows that zero is covered by all 95% Confidence Interval (CI) estimates for coefficients of *citizenp* across the demographic groups.

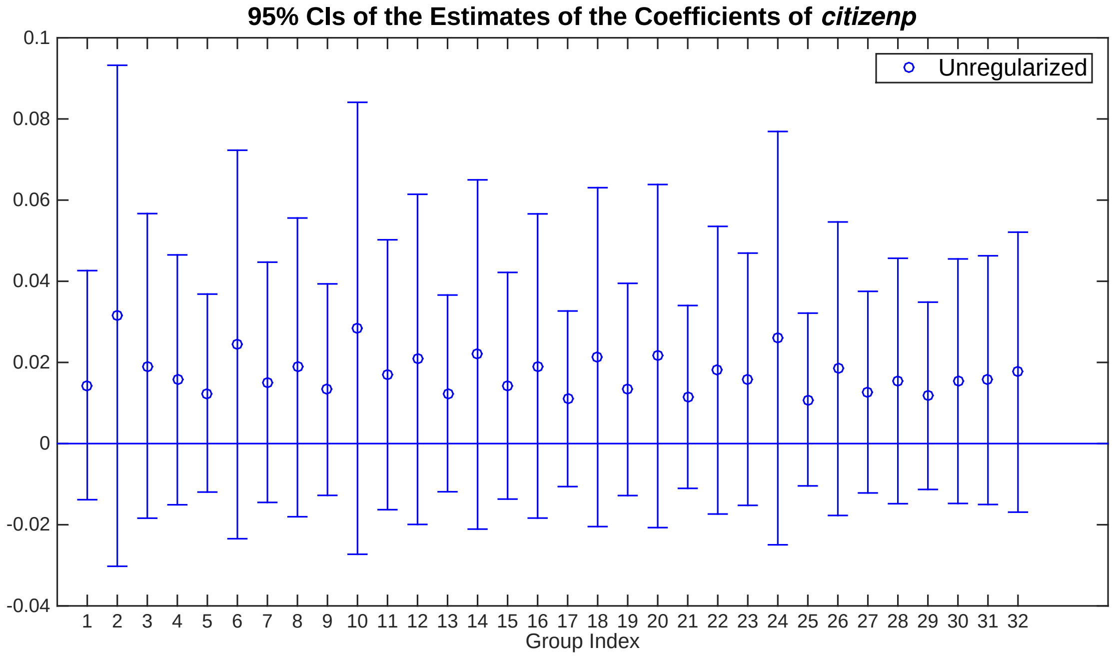

Figure 1: The unregularized estimates for an irrelevant regressor of *citizenp*

Second, the other 21 irrelevant regressors identified by regularized method normally have significant impacts on BMI for minority of groups, estimated by unregularized method. But these regressors cannot be theoretically identified as irrelevant regressors by unregularized method. For example, *rg_mw* has significant impacts on BMI for 13 demographic groups, and the number is higher than that of any other 23 irrelevant regressors identified by regularized method. These results further confirm that the irrelevant regressors identified by the variable selection procedure are truly irrelevant.

4.2.2 Varying Impacts

To further examine the effects of relevant regressors on BMI, we conduct a post-selection estimation using the unregularized estimator for the varying-coefficient model only including the relevant regressors (i.e. equation (2.11)). The full estimated results are reported in Appendix

Table 5: No of groups with significant impacts from corresponding determinants on BMI estimated by unregularized estimator without variable selection

lifestyle factors		socio-economic factors		other factors	
Variable	No of Groups	Variable	No of Groups	Variable	No of Groups
vig_l1*	6	educ1	31	married	18
vig_l2	16	occup1	24	us_born	27
vig_l3	21	occup2	10	us_m15	17
mod_l1*	3	occup3*	8	us_m5l15*	4
mod_l2*	1	occup4*	5	citizenp*	0
mod_l3	9	working*	8	mental	22
str_l1*	11	unemp*	0	rg_ne*	8
str_l2	20	nowork*	8	rg_mw*	13
str_l3	15	wrkhrs*	12	rg_sth	22
smk_ed*	6	income	23		
smk_sd	2	houseown*	6		
smk_f	12	notcov*	4		
cigsday	10	hp	19		
alc1yr	10	hce_l2*	1		
alc_life	16	hce_l3*	1		
alc_c1	2	hce_l4*	6		
alc_c2	11	hce_l5*	3		
alc_c3	22	hce_l6*	6		
alc_c4	21				
cpuse_1*	0				
cpuse_2*	4				

* indicate irrelevant variable identified by our variable selection method

D1 (for lifestyle factors) and D2 (for socio-economic and other factors) for the sake of space limitation. Generally speaking these estimated coefficients confirm that the selected variables are truly relevant to BMI. Because none of these regressors have their effects over all 32 groups to be constant zero, given zero is not consistently covered by the, at least 95%, CIs[5] of the 32 varying-effects of each regressor.

Taking the regressor of *us_born* as an example, its varying effects on BMI cross 32 demographic groups are shown in Figure 2. For comparison, the figure also shows its varying effects estimated by the unregularized method without model selection. The demographic groups are indicated in the horizontal axis (for details, see Table 2). "×" represents the point estimate from the post-selection estimation, and the red bold vertical line represents the 95% CI estimate. Correspondingly, "◦" and the blue normal vertical line represent the point estimate and 95% CI estimate from the unregularized method. Three results emerge from this figure. First, the post-selection results show that the estimated effects of *us_born* on BMI are positive for all groups and none of the CI's cover zero, which confirms that the regressor of *us_born* is truly relevant to BMI. Second, the CI estimations from the unregularized method are much broader than those from the post-selection method, and most of the CI's from the unregularized method cover the

[5]We cannot obtain CI's for the estimates provided in (2.9). However, after the procedure of variable selection, we are able to calculate the 95% CIs for the post-selection estimates. See Theorem 2 and the discussions under Theorem 4 of Li et al. (2013) for details.

corresponding post-selection estimation, which, to some extent, indicates that the estimation efficiency of the unregularized method is substantially degraded by the irrelevant regressors (Wang and Xia, 2009). Third, effects of *us_born* on BMI are apparently varying across the 32 demographic groups. In particular, the effects are higher for males (groups 1-16) than females (groups 17-32) when age and race are the same, i.e. group 1 vs group 17, 2 vs 18, and so forth. Furthermore, the differences are more significant for Asian groups. As shown in Figure 2, there is almost no overlap between the two corresponding CI estimates, i.e. group 3 vs group 19, 7 vs 23, 11 vs 27, and 15 vs 31. Comparing across groups having the same gender and age range, *us_born* normally has higher impacts for Asian people. Taking the four youngest male groups as an example, being born in US increases BMI by 12.78% for Asians, which is higher than the increases of 6.11%, 11.24%, and 8.69% for white, black and all other races, respectively.

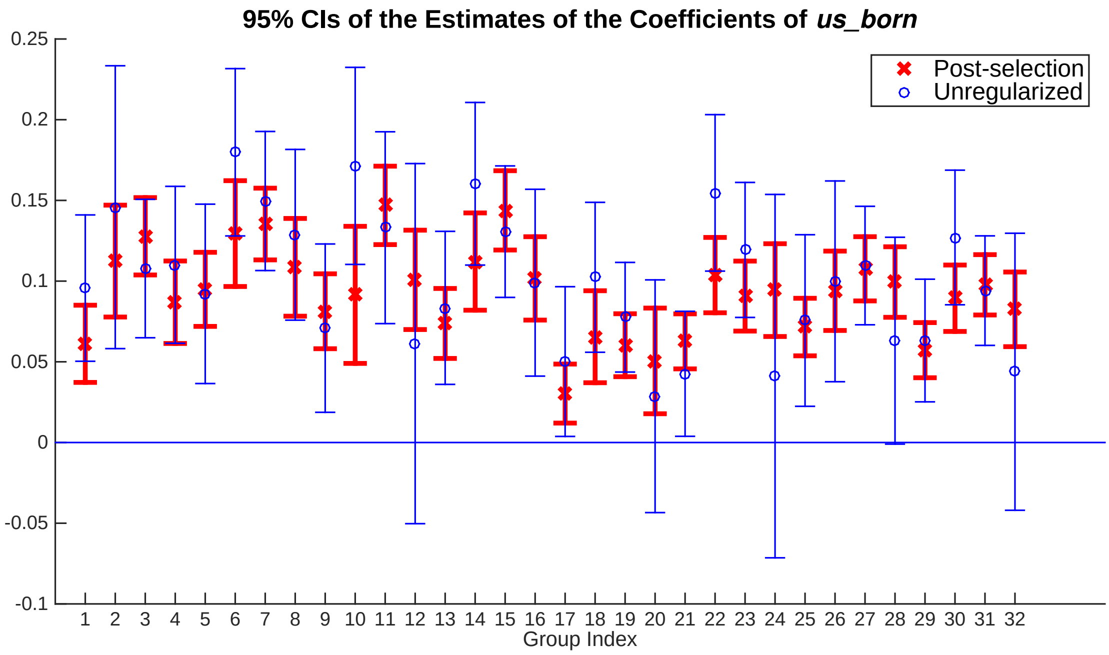

Figure 2: The post-selection estimates and unregularized estimates for a relevant regressor of *us_born*

To further investigate the varying impacts of BMI determinants, we take *str_l3* (i.e. doing strength activities more than three times per week) as another example. Figure 3 shows the point estimates and 95% CI estimates of its impacts on BMI for the 32 demographic groups. Firstly, it is shown that *str_l3* has significant negative impacts on BMI for most males but very few females. In particular, 12 out of 16 male groups have the corresponding 95% CI's below zero; however all female groups except one (i.e. white female aged from 45 to 65) have their corresponding 95% CI's containing zero. Secondly, BMI of black males seems to be more sensitive to *str_l3*, compared to other male groups. For example, doing strength activities more than three times per week can reduce BMI by 8.81% for black males aged from 45 to 65, which is the highest impact of *str_l3* across all groups. Black males also enjoy higher impact of *str_l3* within both the youngest and oldest male groups, compared to those white, Asian or other ethnic males. Thirdly, *str_l3* generally has stronger impacts on BMI for the older males than younger males. For example, doing strength activities more than three times per week reduces BMI by 5.61% for white males aged from 45 to 65, and 5.22% for white males older than 65; while this type of exercise reduces BMI by 2.76% and 4.15% for white males younger than 25 and aged from 25 to 45, respectively. Similar trend can be seen for the other race groups. Therefore according to these estimates, promotions on more frequent strength activities are more efficient to reduce BMI for males than for females, more efficient for older individuals than for younger ones, and more efficient for black ethnicity than for all other ethnicities.

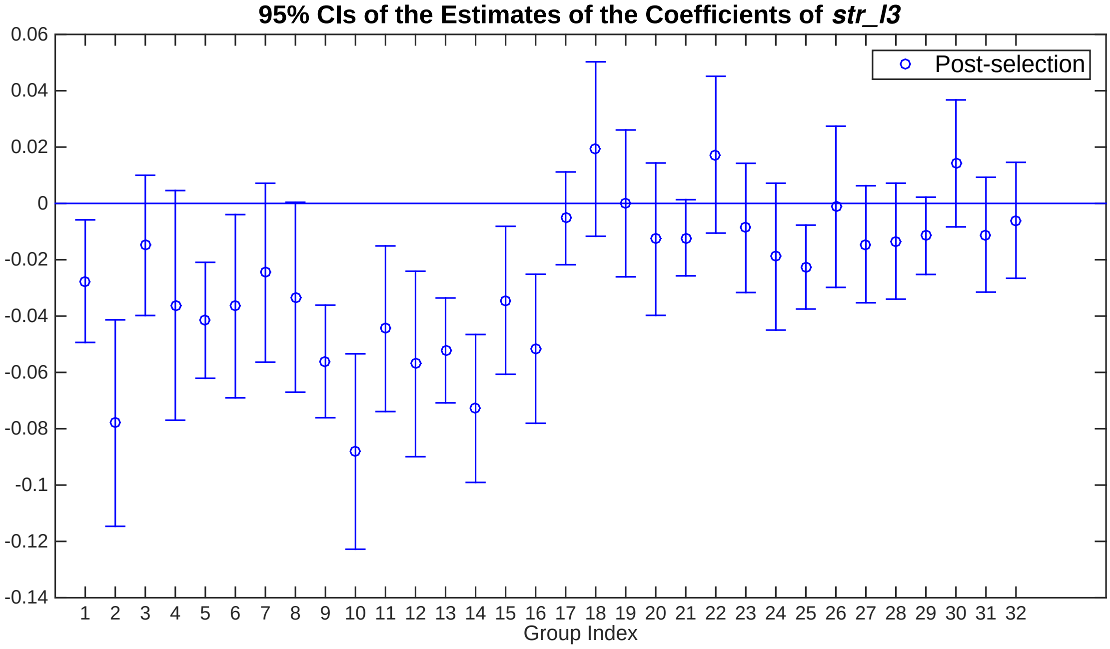

Figure 3: The post-selection estimates for a relevant regressor of *str_l3*

We acknowledge the limitations of this empirical study. Firstly, it is limited by information available in the data set. For example, besides the factors considered in this study, energy intake and dietary habit are important factor for BMI and obesity (see, for example, Hill and Peters (1998) and Millimet and Tchernis (2013)). But unfortunately such information is not available in our data. Secondly, some potential factors, like income, are endogenous and we cannot capture the causal impacts through our modelling method. For example, there are unobserved factors, such as preferences for long-term investments, simultaneously correlated with BMI and income (Kline and Tobias, 2008; Chan and Tobias, 2015). In this empirical study we do not take into account such endogeneity issue, which is out of scope of our study.

5 Conclusions and Discussion

In this paper, we have proposed a variable selection procedure for the categorical varying-coefficient model (Li et al., 2013). We have not only established that the true model (relevant regressors) can be appropriately selected but also established an asymptotic theory for our estimator as a consequence of the oracle property. The finite sample properties of our estimator have been demonstrated through a simulated data example.

In the empirical study, we examine the impacts of a wide range of potential factors proposed in the huge literature on BMI and obesity by using data from the 2013 NHIS of the United States. Specifically, we investigate the varying effects of lifestyle factors, socio-economic factors and other factors (including migrant status, marital status, geographic factors and mental health problems) on BMI over demographic groups (classified by gender, age and ethnicity). Variable selection results from our method have been verified by comparing results on estimated coefficients from the post-selection estimation with those from the unregularized method. In particular, firstly, the post-selection estimates on relevant regressors' impacts on BMI are significantly different from zero. Secondly, CI estimates of relevant regressors' coefficients from the unregularized method are substantially broader than those from our method. Thirdly, most of the estimated coefficients of irrelevant regressors from the unregularized method are not significantly different from zero. Last but not least, the post-selection estimation quantifies the varying impacts of relevant regressors on BMI, which provides valuable guidance to policy makers for more efficient interventions.

We now briefly discuss the case where the cardinality of $\mathcal{D}$ is infinite. Suppose that $\bar{r} = r = 1$ for simplicity. Let $Z_i \in \{0, 1, 2, \ldots, \nu(N) - 1\}$, where $\nu(N) \to \infty$ and $\nu(N)/N \to \omega_3$ for $0 \leq \omega_3 \leq 1$ as $N \to \infty$. In this case, the following model can be considered.

$$Y_i = X_i'\beta_0(Z_i/\nu(N)) + \varepsilon_i, \quad i = 1, \ldots, N \text{ and } t = 1, \ldots, T. \tag{5.1}$$

If we treat $\beta_0(\cdot)$ as a function with a continuous covariate, (5.1) then reduces to the model considered in Wang and Xia (2009), where the variable selection procedure has been fully investigated. This normalization technique is the same as the one employed by Chen et al. (2012) in dealing with time varying-coefficient models, wherein more relevant references can be found.

Appendix A: Estimation Procedure

The estimation procedure for implementing our method is described below.

Steps:

1. Minimize the cross-validation criterion function (2.5) in order to choose $\hat{\Theta}$.

2. Select $\tilde{\gamma}$ defined in (2.12) from a sufficient large set, say $[1, \sqrt[4]{N}]$ by using grid search. For each choice of $\tilde{\gamma}$, implement the estimation proposed by (2.9) in a similar procedure proposed in Wang and Xia (2009). Define

$$\hat{B}_{\tilde{\gamma}}^{(n)} = (\hat{\beta}_{\tilde{\gamma},1}^{(n)}, \ldots, \hat{\beta}_{\tilde{\gamma},m}^{(n)})' = (\hat{b}_{\tilde{\gamma},1}^{(n)}, \ldots, \hat{b}_{\tilde{\gamma},p}^{(n)}) \tag{A.1}$$

to be the estimate obtained in the nth iteration. Then the loss function given above can be locally approximated by

$$\begin{aligned} &\sum_{j=1}^{m}\sum_{i=1}^{N}\left(Y_i - X_i'\beta_j\right)^2 L(Z_i, z^j, \hat{\Theta}) + \sum_{s=1}^{p}\tilde{\gamma}_s \frac{\|b_s\|^2}{\|\hat{b}_{\tilde{\gamma},s}^{(n)}\|} \\ =&\sum_{j=1}^{m}\left(\sum_{i=1}^{N}\left(Y_i - X_i'\beta_j\right)^2 L(Z_i, z^j, \hat{\Theta}) + \sum_{s=1}^{p}\tilde{\gamma}_s \frac{\beta_{j,s}^2}{\|\hat{b}_{\tilde{\gamma},s}^{(n)}\|}\right). \end{aligned} \tag{A.2}$$

The minimizer of (A.2) is provided by $\hat{B}_{\tilde{\gamma}}^{(n+1)} = (\hat{\beta}_{\tilde{\gamma},1}^{(n+1)}, \ldots, \hat{\beta}_{\tilde{\gamma},m}^{(n+1)})'$, where for $j = 1, \ldots, m$

$$\hat{\beta}_{\tilde{\gamma},j}^{(n+1)} = \left(\sum_{i=1}^{N} X_i X_i' L(Z_i, z^j, \hat{\Theta}) + D^{(n)}\right)^{-1} \sum_{i=1}^{N} X_i Y_i L(Z_i, z^j, \hat{\Theta}), \tag{A.3}$$

and $D^{(n)} = \operatorname{diag}\left(\|\hat{b}_{\tilde{\gamma},1}^{(n)}\|^{-1}\gamma_1, \ldots, \|\hat{b}_{\gamma,p}^{(n)}\|^{-1}\gamma_p\right)$. Repeat this procedure until $\|\hat{B}_{\tilde{\gamma}}^{(n+1)} - \hat{B}_{\tilde{\gamma}}^{(n)}\| <$ *tolerance*, where *tolerance* is a sufficiently small number (say, 10^{-10}).

3. Select the optimal estimator based on the modified BIC-type criterion.

4. After removing the irrelevant regressors, carry on the unregularized estimation as proposed in Li et al. (2013).

Appendix B: Data Descriptions and Summary Statistics

Variable	Definition	Mean	St.D
Y			
BMI	body mass index	27.96	6.01
Z			
sex	0 for female and 1 for male	0.49	0.50
age	0 for age<25, 1 for 25<=age<=44, 2 for 45<=age<=64, and 3 for age>=65	1.39	0.75
race	0 for white, 1 for black, 2 for asian, 3 for all the other races	0.33	0.67
X			
lifestyle factors			
vig_l0	1 if never do vigorous activities, 0 otherwise (reference group)	0.45	0.50
vig_l1	1 if do vigorous activities less than once per week, 0 otherwise	0.04	0.19
vig_l2	1 if do vigorous activities more than one time and less than three times per week, 0 otherwise	0.28	0.45
vig_l3	1 if do vigorous activities more than three times per week, 0 otherwise	0.23	0.42
mod_l0	1 if never do light/moderate activities, 0 otherwise (reference group)	0.35	0.48
mod_l1	1 if do light/moderate activities less than once per week, 0 otherwise	0.02	0.15
mod_l2	1 if do light/moderate activities more than one time and less than three times per week, 0 otherwise	0.29	0.46
mod_l3	1 if do light/moderate activities more than three times per week, 0 otherwise	0.33	0.47
str_l0	1 if never do strength activities, 0 otherwise (reference group)	0.66	0.47
str_l1	1 if do strength activities less than once per week, 0 otherwise	0.02	0.14
str_l2	1 if do strength activities more than one time and less than three times per week, 0 otherwise	0.20	0.40
str_l3	1 if do strength activities more than three times per week, 0 otherwise	0.12	0.32
smk_ed	1 if current every day smoker, 0 otherwise	0.13	0.34
smk_sd	1 if current some day smoker, 0 otherwise	0.04	0.20
smk_f	1 if former smoker, 0 otherwise	0.20	0.40
smk_n	1 if never smoke, 0 otherwise (reference group)	0.62	0.48
cigsday	number of cigarettes per day	1.98	5.52
alc1yr	1 if Ever had 12+ drinks in any one year, 0 otherwise	0.72	0.45
alc_life	1 if Had 12+ drinks in entire life, 0 otherwise	0.13	0.33
alc_c0	1 if do not drink at all currently, 0 otherwise (reference group)	0.26	0.44
alc_c1	1 if current infrequent drinker, 0 otherwise	0.12	0.33
alc_c2	1 if current light drinker, 0 otherwise	0.36	0.48
alc_c3	1 if current moderate drinker, 0 otherwise	0.19	0.39
alc_c4	1 if current heavier drinker, 0 otherwise	0.06	0.25
cpuse_0	1 if never or almost never use computer, 0 otherwise (reference group)	0.15	0.35
cpuse_1	1 if use computer for some/most days, 0 otherwise	0.18	0.38
cpuse_2	1 if use computer on every day, 0 otherwise	0.67	0.47
socio-economic factors			
educ1	number of years of school completed	15.54	3.08
occup1	1 if management, business, science, and arts occupations, 0 otherwise	0.38	0.49
occup2	1 if service occupations, 0 otherwise	0.18	0.38
occup3	1 if sales and office occupations, 0 otherwise	0.23	0.42

occup4	1 if natural resources, construction, and maintenance occupations, 0 otherwise	0.09	0.29
occup5	1 if production, transportation, and material moving occupations, 0 otherwise (reference group)	0.12	0.33
working	1 if working or with job last week, 0 otherwise	0.88	0.32
unemp	1 if looking for job last week, 0 otherwise	0.05	0.21
nowork	1 if not working at a job last week, 0 otherwise	0.05	0.22
retired	1 if retired, 0 otherwise (reference group)	0.02	0.15
wrkhrs	hours worked last week	35.46	17.28
lnincome	nature logrithm of total earnings last year	10.20	0.94
houseown	1 if own or being bought the house, 0 otherwise	0.56	0.50
notcov	1 if not have health insurance coverage, 0 otherwise	0.20	0.40
hp	1 if ever seen/talked to health professional in the last 12 months, 0 otherwise	0.79	0.40
hce_l1	1 if amount family spent for medical care is 0, 0 otherwise (reference group)	0.13	0.33
hce_l2	1 if amount family spent for medical care is less than $500 but more than 0, 0 otherwise	0.37	0.48
hce_l3	1 if amount family spent for medical care is less than $1999 but more than $500, 0 otherwise	0.30	0.46
hce_l4	1 if amount family spent for medical care is less than $2999 but more than $2000, 0 otherwise	0.09	0.29
hce_l5	1 if amount family spent for medical care is less than $4999 but more than $3000, 0 otherwise	0.06	0.24
hce_l6	1 if amount family spent for medical care is $5000 or more, 0 otherwise	0.06	0.23
other factors			
married	1 if married or de facto, 0 otherwise	0.51	0.50
us_born	1 if born in the US, 0 otherwise	0.81	0.39
us_m15	1 if stay in the US for more than 15 years, 0 otherwise	0.12	0.32
us_m5l15	1 if stay in the US for more than 5 years but less than 15 years, 0 otherwise	0.06	0.24
us_l5	1 if stay in the US for less than 5 years, 0 otherwise (reference group)	0.02	0.12
citizenp	1 if U.S. citizen, 0 otherwise	0.90	0.30
mental	1 if have depression/anxiety/emotional problem, 0 otherwise	0.01	0.12
rg_ne	1 if live in north east, 0 otherwise	0.16	0.37
rg_mw	1 if live in midwest, 0 otherwise	0.21	0.41
rg_sth	1 if live in south, 0 otherwise	0.36	0.48
rg_west	1 if live in west, 0 otherwise (reference group)	0.27	0.44

Appendix C1: Estimated parameters of lifestyle factors by the unregularized estimation

Group Index	vig_l1	vig_l2	vig_l3	mod_l1	mod_l2	mod_l3	str_l1	str_l2	str_l3	smk_ed	smk_sd	smk_f	cigsday	alc1yr	alc_life	alc_c1	alc_c2	alc_c3	alc_c4	cpuse_1	cpuse_2
1	-0.0435	-0.0218	-0.0446	0.0187	0.0115	-0.0109	-0.0468	-0.0335	-0.0290	0.0050	0.0193	0.0219	-0.0006	0.0169	0.0350	0.0056	-0.0270	-0.0526	-0.0556	-0.0081	-0.0024
	(0.0164)	(0.0077)	(0.0085)	(0.0191)	(0.0081)	(0.0074)	(0.0171)	(0.0087)	(0.0104)	(0.0182)	(0.0175)	(0.0070)	(0.0012)	(0.0143)	(0.0140)	(0.0119)	(0.0121)	(0.0132)	(0.0149)	(0.0127)	(0.0111)
2	-0.0278	0.0127	-0.0024	-0.0458	-0.0182	-0.0012	-0.0477	-0.0349	-0.0763	-0.0069	0.0012	0.0206	-0.0014	0.0373	0.0789	-0.0373	-0.0202	-0.0382	-0.0560	-0.0017	-0.0208
	(0.0253)	(0.0148)	(0.0154)	(0.0334)	(0.0138)	(0.0146)	(0.0266)	(0.0173)	(0.0159)	(0.0310)	(0.0264)	(0.0124)	(0.0019)	(0.0268)	(0.0195)	(0.0230)	(0.0271)	(0.0263)	(0.0307)	(0.0194)	(0.0192)
3	0.0466	-0.0114	-0.0130	0.0163	0.0044	-0.0066	-0.0307	-0.0193	-0.0167	0.0471	0.0253	0.0340	-0.0021	0.0378	0.0215	-0.0054	-0.0413	-0.0688	-0.1036	-0.0132	-0.0136
	(0.0346)	(0.0103)	(0.0104)	(0.0267)	(0.0112)	(0.0095)	(0.0412)	(0.0131)	(0.0116)	(0.0195)	(0.0250)	(0.0131)	(0.0013)	(0.0160)	(0.0154)	(0.0150)	(0.0150)	(0.0151)	(0.0226)	(0.0131)	(0.0123)
4	-0.0287	-0.0080	-0.0225	0.0161	-0.0031	-0.0137	-0.0426	-0.0400	-0.0347	0.0170	0.0297	0.0222	-0.0015	0.0360	0.0382	-0.0097	-0.0359	-0.0658	-0.0716	-0.0063	-0.0188
	(0.0153)	(0.0096)	(0.0103)	(0.0173)	(0.0094)	(0.0078)	(0.0214)	(0.0104)	(0.0176)	(0.0173)	(0.0170)	(0.0094)	(0.0010)	(0.0167)	(0.0158)	(0.0150)	(0.0161)	(0.0152)	(0.0176)	(0.0131)	(0.0125)
5	-0.0347	-0.0202	-0.0325	0.0157	-0.0008	-0.0275	-0.0493	-0.0417	-0.0425	-0.0081	0.0224	0.0198	-0.0005	0.0089	0.0223	0.0055	-0.0308	-0.0661	-0.0719	0.0160	0.0249
	(0.0150)	(0.0073)	(0.0089)	(0.0179)	(0.0078)	(0.0068)	(0.0193)	(0.0082)	(0.0111)	(0.0161)	(0.0149)	(0.0079)	(0.0011)	(0.0158)	(0.0151)	(0.0134)	(0.0142)	(0.0152)	(0.0159)	(0.0109)	(0.0103)
6	-0.0333	0.0124	-0.0261	0.0065	0.0154	0.0000	-0.0711	-0.0348	-0.0421	-0.0068	-0.0132	0.0115	-0.0021	0.0106	0.0387	-0.0083	-0.0155	-0.0402	-0.0278	0.0250	0.0025
	(0.0331)	(0.0148)	(0.0147)	(0.0270)	(0.0139)	(0.0145)	(0.0357)	(0.0158)	(0.0179)	(0.0234)	(0.0209)	(0.0138)	(0.0016)	(0.0233)	(0.0209)	(0.0219)	(0.0224)	(0.0236)	(0.0289)	(0.0186)	(0.0164)
7	0.0449	-0.0140	-0.0073	0.0019	-0.0201	-0.0099	-0.0493	-0.0096	-0.0242	0.0388	0.0167	0.0237	-0.0025	0.0310	0.0097	0.0084	-0.0362	-0.0586	-0.0791	0.0078	0.0072
	(0.0234)	(0.0118)	(0.0119)	(0.0395)	(0.0107)	(0.0116)	(0.0167)	(0.0114)	(0.0143)	(0.0198)	(0.0237)	(0.0132)	(0.0017)	(0.0204)	(0.0202)	(0.0173)	(0.0177)	(0.0192)	(0.0230)	(0.0146)	(0.0145)
8	-0.0331	-0.0143	-0.0251	0.0041	-0.0044	-0.0191	-0.0360	-0.0392	-0.0316	0.0207	0.0376	0.0233	-0.0018	0.0282	0.0182	-0.0057	-0.0445	-0.0806	-0.0765	0.0114	-0.0083
	(0.0156)	(0.0111)	(0.0109)	(0.0150)	(0.0107)	(0.0091)	(0.0363)	(0.0121)	(0.0149)	(0.0185)	(0.0232)	(0.0107)	(0.0012)	(0.0214)	(0.0184)	(0.0181)	(0.0213)	(0.0202)	(0.0217)	(0.0138)	(0.0141)
9	-0.0156	-0.0278	-0.0424	0.0018	0.0126	-0.0179	-0.0655	-0.0501	-0.0576	-0.0043	0.0234	0.0131	-0.0019	0.0118	0.0083	0.0179	-0.0279	-0.0571	-0.0655	0.0054	0.0203
	(0.0193)	(0.0069)	(0.0078)	(0.0213)	(0.0080)	(0.0068)	(0.0240)	(0.0083)	(0.0096)	(0.0165)	(0.0161)	(0.0068)	(0.0009)	(0.0125)	(0.0119)	(0.0112)	(0.0117)	(0.0126)	(0.0134)	(0.0101)	(0.0096)
10	-0.0369	-0.0241	-0.0203	0.0114	0.0023	-0.0024	-0.0510	-0.0342	-0.0882	-0.0092	0.0165	0.0097	-0.0028	-0.0175	0.0110	-0.0134	-0.0037	-0.0230	0.0149	-0.0025	-0.0033
	(0.0272)	(0.0120)	(0.0130)	(0.0280)	(0.0121)	(0.0117)	(0.0345)	(0.0131)	(0.0171)	(0.0255)	(0.0281)	(0.0143)	(0.0016)	(0.0190)	(0.0162)	(0.0165)	(0.0169)	(0.0219)	(0.0268)	(0.0147)	(0.0152)
11	0.0271	-0.0183	-0.0280	0.0297	0.0095	-0.0063	-0.0426	-0.0347	-0.0434	0.0343	0.0474	0.0181	-0.0036	0.0086	0.0114	0.0021	-0.0299	-0.0594	-0.0517	0.0040	0.0059
	(0.0219)	(0.0105)	(0.0101)	(0.0304)	(0.0100)	(0.0110)	(0.0260)	(0.0115)	(0.0144)	(0.0249)	(0.0216)	(0.0112)	(0.0013)	(0.0201)	(0.0175)	(0.0175)	(0.0172)	(0.0181)	(0.0241)	(0.0147)	(0.0148)
12	-0.0179	-0.0150	-0.0361	-0.0009	0.0098	-0.0124	-0.0605	-0.0528	-0.0608	-0.0104	0.0358	0.0050	-0.0021	-0.0034	0.0121	0.0050	-0.0182	-0.0454	-0.0384	0.0059	-0.0024
	(0.0170)	(0.0110)	(0.0113)	(0.0183)	(0.0114)	(0.0095)	(0.0231)	(0.0110)	(0.0159)	(0.0168)	(0.0198)	(0.0111)	(0.0010)	(0.0203)	(0.0175)	(0.0190)	(0.0203)	(0.0210)	(0.0265)	(0.0119)	(0.0114)
13	-0.0326	-0.0193	-0.0313	-0.0026	0.0069	-0.0262	-0.0557	-0.0406	-0.0530	-0.0213	0.0206	0.0170	-0.0004	0.0086	0.0353	0.0153	-0.0269	-0.0635	-0.0712	0.0147	0.0295
	(0.0140)	(0.0073)	(0.0085)	(0.0218)	(0.0078)	(0.0068)	(0.0169)	(0.0077)	(0.0095)	(0.0143)	(0.0132)	(0.0070)	(0.0010)	(0.0141)	(0.0130)	(0.0126)	(0.0133)	(0.0141)	(0.0138)	(0.0108)	(0.0097)
14	-0.0563	-0.0108	-0.0245	-0.0020	0.0016	-0.0001	-0.0338	-0.0321	-0.0725	0.0033	0.0192	0.0276	-0.0029	0.0007	0.0392	-0.0298	-0.0180	-0.0372	-0.0348	0.0055	0.0014
	(0.0206)	(0.0115)	(0.0124)	(0.0290)	(0.0118)	(0.0112)	(0.0292)	(0.0132)	(0.0126)	(0.0222)	(0.0217)	(0.0127)	(0.0014)	(0.0197)	(0.0154)	(0.0176)	(0.0190)	(0.0212)	(0.0228)	(0.0151)	(0.0143)
15	0.0321	-0.0141	-0.0138	0.0039	-0.0080	-0.0130	-0.0429	-0.0244	-0.0333	0.0291	0.0298	0.0288	-0.0020	0.0179	0.0054	0.0137	-0.0285	-0.0585	-0.0769	0.0251	0.0283
	(0.0186)	(0.0092)	(0.0092)	(0.0259)	(0.0102)	(0.0087)	(0.0200)	(0.0099)	(0.0128)	(0.0176)	(0.0187)	(0.0096)	(0.0012)	(0.0169)	(0.0196)	(0.0165)	(0.0135)	(0.0150)	(0.0185)	(0.0155)	(0.0144)
16	-0.0498	-0.0092	-0.0222	-0.0034	-0.0047	-0.0218	-0.0431	-0.0347	-0.0503	-0.0112	0.0410	0.0177	-0.0012	0.0034	0.0330	0.0019	-0.0263	-0.0613	-0.0610	0.0145	0.0114
	(0.0193)	(0.0089)	(0.0095)	(0.0169)	(0.0096)	(0.0078)	(0.0247)	(0.0109)	(0.0116)	(0.0164)	(0.0173)	(0.0090)	(0.0010)	(0.0165)	(0.0162)	(0.0163)	(0.0166)	(0.0168)	(0.0182)	(0.0110)	(0.0108)
17	-0.0262	-0.0206	-0.0383	0.0622	0.0045	-0.0060	-0.0071	-0.0091	-0.0048	-0.0482	-0.0171	0.0199	0.0003	0.0211	0.0326	0.0109	-0.0065	-0.0172	-0.0251	-0.0057	-0.0146
	(0.0159)	(0.0066)	(0.0075)	(0.0204)	(0.0071)	(0.0064)	(0.0246)	(0.0067)	(0.0081)	(0.0154)	(0.0111)	(0.0060)	(0.0008)	(0.0125)	(0.0125)	(0.0107)	(0.0104)	(0.0108)	(0.0135)	(0.0086)	(0.0079)

18	-0.0147	-0.0308	-0.0414	0.0310	0.0137	-0.0040	0.0219	-0.0072	0.0190	-0.0371	0.0306	0.0196	-0.0022	0.0263	0.0534	-0.0269	-0.0191	-0.0185	-0.0195	0.0049	0.0003
	(0.0218)	(0.0110)	(0.0124)	(0.0244)	(0.0112)	(0.0105)	(0.0239)	(0.0119)	(0.0139)	(0.0208)	(0.0203)	(0.0092)	(0.0012)	(0.0200)	(0.0153)	(0.0161)	(0.0169)	(0.0173)	(0.0198)	(0.0124)	(0.0131)
19	0.0026	-0.0135	-0.0166	0.0419	0.0156	-0.0035	-0.0195	-0.0055	-0.0013	-0.0312	0.0137	0.0151	0.0006	0.0437	0.0168	-0.0139	-0.0368	-0.0383	-0.0726	0.0040	-0.0023
	(0.0138)	(0.0083)	(0.0124)	(0.0177)	(0.0107)	(0.0110)	(0.0159)	(0.0092)	(0.0137)	(0.0163)	(0.0218)	(0.0075)	(0.0013)	(0.0148)	(0.0133)	(0.0130)	(0.0119)	(0.0137)	(0.0225)	(0.0106)	(0.0107)
20	-0.0251	-0.0181	-0.0199	0.0393	-0.0008	-0.0047	-0.0170	-0.0182	-0.0130	-0.0031	0.0036	0.0107	-0.0026	0.0543	0.0424	-0.0152	-0.0380	-0.0487	-0.0604	-0.0131	-0.0221
	(0.0150)	(0.0087)	(0.0114)	(0.0178)	(0.0096)	(0.0099)	(0.0202)	(0.0096)	(0.0147)	(0.0227)	(0.0166)	(0.0083)	(0.0009)	(0.0169)	(0.0177)	(0.0199)	(0.0148)	(0.0152)	(0.0185)	(0.0169)	(0.0156)
21	-0.0096	-0.0170	-0.0360	0.0252	-0.0033	-0.0126	-0.0214	-0.0159	-0.0117	-0.0230	-0.0024	0.0204	-0.0006	0.0286	0.0189	0.0113	-0.0137	-0.0313	-0.0468	0.0001	-0.0018
	(0.0138)	(0.0066)	(0.0073)	(0.0164)	(0.0064)	(0.0065)	(0.0170)	(0.0062)	(0.0074)	(0.0126)	(0.0096)	(0.0066)	(0.0007)	(0.0126)	(0.0118)	(0.0110)	(0.0102)	(0.0106)	(0.0125)	(0.0092)	(0.0081)
22	-0.0127	-0.0174	-0.0338	0.0245	0.0180	-0.0011	-0.0128	-0.0147	0.0140	-0.0156	0.0268	0.0141	-0.0018	0.0053	0.0518	-0.0195	-0.0138	-0.0182	0.0020	0.0269	0.0277
	(0.0263)	(0.0098)	(0.0126)	(0.0305)	(0.0105)	(0.0100)	(0.0280)	(0.0110)	(0.0127)	(0.0195)	(0.0292)	(0.0103)	(0.0011)	(0.0193)	(0.0170)	(0.0192)	(0.0176)	(0.0179)	(0.0228)	(0.0139)	(0.0135)
23	0.0124	-0.0081	-0.0116	-0.0007	0.0006	-0.0105	-0.0223	-0.0120	-0.0097	-0.0291	0.0029	0.0064	0.0004	0.0452	0.0329	0.0033	-0.0206	-0.0388	-0.0542	0.0162	0.0091
	(0.0162)	(0.0091)	(0.0122)	(0.0213)	(0.0095)	(0.0098)	(0.0202)	(0.0101)	(0.0113)	(0.0211)	(0.0205)	(0.0095)	(0.0014)	(0.0195)	(0.0156)	(0.0170)	(0.0177)	(0.0194)	(0.0194)	(0.0125)	(0.0122)
24	-0.0161	-0.0153	-0.0195	-0.0135	-0.0215	-0.0170	-0.0483	-0.0337	-0.0197	0.0079	0.0036	0.0086	-0.0028	0.0536	0.0451	-0.0279	-0.0271	-0.0537	-0.0434	-0.0174	-0.0020
	(0.0176)	(0.0118)	(0.0126)	(0.0208)	(0.0118)	(0.0109)	(0.0224)	(0.0104)	(0.0131)	(0.0179)	(0.0160)	(0.0107)	(0.0008)	(0.0186)	(0.0168)	(0.0174)	(0.0175)	(0.0178)	(0.0235)	(0.0163)	(0.0168)
25	-0.0130	-0.0145	-0.0280	0.0230	-0.0014	-0.0164	-0.0312	-0.0266	-0.0234	-0.0477	-0.0056	0.0081	-0.0005	0.0004	0.0133	0.0211	0.0048	-0.0152	-0.0321	-0.0074	-0.0039
	(0.0119)	(0.0056)	(0.0065)	(0.0173)	(0.0061)	(0.0057)	(0.0198)	(0.0068)	(0.0072)	(0.0119)	(0.0114)	(0.0061)	(0.0007)	(0.0119)	(0.0114)	(0.0098)	(0.0093)	(0.0096)	(0.0102)	(0.0081)	(0.0076)
26	-0.0116	-0.0448	-0.0355	-0.0104	0.0154	0.0023	0.0022	-0.0131	0.0003	-0.0422	-0.0197	0.0130	-0.0028	0.0183	0.0421	-0.0042	-0.0048	-0.0047	-0.0134	0.0122	-0.0023
	(0.0224)	(0.0114)	(0.0131)	(0.0223)	(0.0099)	(0.0114)	(0.0224)	(0.0111)	(0.0151)	(0.0238)	(0.0233)	(0.0107)	(0.0012)	(0.0185)	(0.0153)	(0.0166)	(0.0178)	(0.0166)	(0.0281)	(0.0134)	(0.0127)
27	-0.0063	-0.0199	-0.0192	0.0053	0.0118	-0.0059	0.0000	-0.0052	-0.0138	-0.0284	-0.0091	0.0125	-0.0014	0.0300	0.0120	0.0103	-0.0139	-0.0289	-0.0484	0.0053	0.0053
	(0.0153)	(0.0085)	(0.0108)	(0.0154)	(0.0085)	(0.0088)	(0.0224)	(0.0085)	(0.0104)	(0.0161)	(0.0193)	(0.0078)	(0.0009)	(0.0154)	(0.0146)	(0.0139)	(0.0137)	(0.0144)	(0.0154)	(0.0119)	(0.0100)
28	-0.0141	-0.0120	-0.0250	-0.0036	-0.0159	-0.0199	-0.0284	-0.0231	-0.0135	-0.0324	-0.0108	0.0086	-0.0021	0.0150	0.0268	-0.0022	-0.0131	-0.0344	-0.0362	-0.0062	-0.0098
	(0.0119)	(0.0092)	(0.0080)	(0.0140)	(0.0096)	(0.0089)	(0.0164)	(0.0078)	(0.0107)	(0.0122)	(0.0117)	(0.0076)	(0.0006)	(0.0185)	(0.0172)	(0.0150)	(0.0150)	(0.0138)	(0.0165)	(0.0107)	(0.0115)
29	-0.0152	-0.0213	-0.0359	0.0281	-0.0034	-0.0167	-0.0246	-0.0164	-0.0112	-0.0274	-0.0054	0.0110	-0.0009	0.0234	0.0168	0.0314	-0.0116	-0.0226	-0.0373	-0.0073	0.0036
	(0.0114)	(0.0059)	(0.0064)	(0.0149)	(0.0061)	(0.0055)	(0.0152)	(0.0057)	(0.0076)	(0.0107)	(0.0090)	(0.0059)	(0.0006)	(0.0104)	(0.0101)	(0.0103)	(0.0097)	(0.0095)	(0.0117)	(0.0078)	(0.0072)
30	-0.0276	-0.0400	-0.0486	0.0077	0.0090	-0.0040	0.0239	-0.0114	0.0143	-0.0256	0.0181	0.0159	-0.0022	0.0243	0.0504	-0.0160	-0.0181	-0.0222	-0.0296	0.0118	0.0075
	(0.0176)	(0.0090)	(0.0101)	(0.0211)	(0.0084)	(0.0084)	(0.0261)	(0.0084)	(0.0100)	(0.0167)	(0.0205)	(0.0100)	(0.0009)	(0.0173)	(0.0136)	(0.0143)	(0.0173)	(0.0169)	(0.0235)	(0.0116)	(0.0119)
31	0.0102	-0.0208	-0.0245	0.0066	-0.0009	-0.0109	-0.0252	-0.0065	-0.0117	-0.0212	0.0062	0.0106	-0.0007	0.0417	0.0232	0.0103	-0.0238	-0.0318	-0.0546	0.0141	0.0129
	(0.0148)	(0.0069)	(0.0094)	(0.0150)	(0.0073)	(0.0074)	(0.0216)	(0.0077)	(0.0085)	(0.0160)	(0.0169)	(0.0071)	(0.0009)	(0.0147)	(0.0112)	(0.0124)	(0.0143)	(0.0152)	(0.0162)	(0.0101)	(0.0095)
32	-0.0204	-0.0165	-0.0268	0.0024	-0.0199	-0.0227	-0.0347	-0.0238	-0.0077	-0.0187	0.0003	0.0117	-0.0020	0.0343	0.0388	-0.0004	-0.0214	-0.0414	-0.0406	-0.0003	0.0095
	(0.0130)	(0.0083)	(0.0083)	(0.0146)	(0.0083)	(0.0081)	(0.0172)	(0.0071)	(0.0098)	(0.0201)	(0.0125)	(0.0074)	(0.0008)	(0.0144)	(0.0130)	(0.0125)	(0.0122)	(0.0121)	(0.0147)	(0.0106)	(0.0106)

[a] Standard deviation based on 200 bootstrap replications is reported in brackets below the corresponding estimated coefficient.

Appendix C2: Estimated parameters of socio-economic factors by the unregularized estimation

Group Index	educ1	occup1	occup2	occup3	occup4	working	unemp	nowork	wrkhrs	income	houseown	notcov	hp	hce_l2	hce_l3	hce_l4	hce_l5	hce_l6
1	-0.0092	-0.0295	-0.0382	-0.0256	0.0072	-0.0657	0.0044	-0.0517	0.0009	0.0160	0.0007	0.0080	0.0111	0.0026	0.0162	0.0178	0.0256	0.0253
	(0.0012)	(0.0113)	(0.0126)	(0.0114)	(0.0142)	(0.0169)	(0.0242)	(0.0185)	(0.0002)	(0.0041)	(0.0066)	(0.0090)	(0.0085)	(0.0098)	(0.0103)	(0.0137)	(0.0144)	(0.0164)
2	-0.0064	-0.0493	-0.0332	-0.0532	-0.0664	-0.0092	0.0571	-0.0216	0.0006	0.0274	-0.0098	0.0206	0.0019	-0.0024	0.0212	0.0112	0.0425	0.0131
	(0.0027)	(0.0192)	(0.0202)	(0.0203)	(0.0276)	(0.0294)	(0.0427)	(0.0322)	(0.0005)	(0.0078)	(0.0124)	(0.0167)	(0.0185)	(0.0167)	(0.0185)	(0.0225)	(0.0198)	(0.0187)
3	-0.0103	-0.0459	-0.0426	-0.0236	0.0161	-0.0543	-0.0013	-0.0573	0.0006	0.0240	-0.0110	0.0159	0.0226	-0.0010	0.0156	-0.0022	-0.0040	0.0179
	(0.0017)	(0.0138)	(0.0129)	(0.0129)	(0.0173)	(0.0171)	(0.0194)	(0.0199)	(0.0004)	(0.0057)	(0.0083)	(0.0089)	(0.0095)	(0.0123)	(0.0128)	(0.0162)	(0.0176)	(0.0184)
4	-0.0069	-0.0236	-0.0163	-0.0157	0.0231	-0.0662	-0.0108	-0.0742	0.0008	0.0202	0.0030	0.0191	0.0027	-0.0224	-0.0153	-0.0088	-0.0096	-0.0201
	(0.0020)	(0.0140)	(0.0147)	(0.0128)	(0.0171)	(0.0214)	(0.0263)	(0.0241)	(0.0003)	(0.0043)	(0.0077)	(0.0110)	(0.0129)	(0.0132)	(0.0138)	(0.0173)	(0.0171)	(0.0188)
5	-0.0110	-0.0397	-0.0394	-0.0216	0.0138	-0.0396	0.0333	-0.0241	0.0009	0.0018	0.0001	0.0027	0.0108	-0.0147	0.0002	0.0204	0.0188	0.0070
	(0.0011)	(0.0120)	(0.0124)	(0.0128)	(0.0149)	(0.0171)	(0.0212)	(0.0199)	(0.0002)	(0.0041)	(0.0060)	(0.0090)	(0.0074)	(0.0102)	(0.0109)	(0.0145)	(0.0151)	(0.0153)
6	-0.0046	-0.0544	-0.0071	-0.0286	-0.0055	0.0119	0.0381	0.0014	0.0005	0.0087	-0.0186	0.0013	0.0211	0.0056	0.0095	0.0156	0.0508	0.0071
	(0.0026)	(0.0206)	(0.0202)	(0.0207)	(0.0346)	(0.0298)	(0.0354)	(0.0319)	(0.0005)	(0.0068)	(0.0119)	(0.0139)	(0.0148)	(0.0131)	(0.0163)	(0.0184)	(0.0217)	(0.0209)
7	-0.0086	-0.0418	-0.0306	-0.0148	0.0630	-0.0113	-0.0122	-0.0342	0.0000	0.0096	-0.0192	0.0019	0.0339	-0.0240	-0.0081	0.0020	-0.0082	-0.0169
	(0.0021)	(0.0174)	(0.0161)	(0.0168)	(0.0219)	(0.0170)	(0.0236)	(0.0220)	(0.0003)	(0.0054)	(0.0091)	(0.0112)	(0.0112)	(0.0140)	(0.0155)	(0.0198)	(0.0216)	(0.0200)
8	-0.0076	-0.0234	-0.0256	0.0015	0.0499	-0.0240	0.0022	-0.0528	0.0004	0.0056	0.0059	-0.0033	0.0038	-0.0158	-0.0143	0.0069	-0.0078	-0.0209
	(0.0024)	(0.0210)	(0.0191)	(0.0193)	(0.0263)	(0.0234)	(0.0277)	(0.0274)	(0.0003)	(0.0051)	(0.0088)	(0.0127)	(0.0160)	(0.0139)	(0.0138)	(0.0196)	(0.0181)	(0.0176)
9	-0.0080	-0.0211	-0.0171	-0.0191	0.0256	-0.0393	0.0161	-0.0295	0.0007	0.0040	-0.0096	0.0050	0.0222	-0.0115	-0.0002	0.0088	0.0188	0.0123
	(0.0011)	(0.0114)	(0.0114)	(0.0108)	(0.0147)	(0.0186)	(0.0219)	(0.0199)	(0.0002)	(0.0040)	(0.0068)	(0.0093)	(0.0081)	(0.0100)	(0.0105)	(0.0126)	(0.0145)	(0.0154)
10	-0.0056	-0.0279	-0.0139	-0.0303	-0.0381	0.0337	0.0323	0.0209	-0.0001	0.0206	-0.0154	0.0123	0.0220	0.0084	0.0120	0.0343	0.0362	0.0248
	(0.0020)	(0.0185)	(0.0172)	(0.0174)	(0.0225)	(0.0343)	(0.0381)	(0.0379)	(0.0004)	(0.0060)	(0.0123)	(0.0146)	(0.0172)	(0.0144)	(0.0156)	(0.0203)	(0.0199)	(0.0193)
11	-0.0081	-0.0332	-0.0200	-0.0372	0.0014	-0.0261	-0.0251	-0.0293	0.0002	0.0154	-0.0251	0.0041	0.0372	-0.0218	-0.0164	-0.0128	-0.0348	0.0093
	(0.0018)	(0.0164)	(0.0155)	(0.0149)	(0.0240)	(0.0180)	(0.0232)	(0.0224)	(0.0003)	(0.0055)	(0.0088)	(0.0117)	(0.0124)	(0.0134)	(0.0144)	(0.0169)	(0.0186)	(0.0267)
12	-0.0079	-0.0225	0.0108	-0.0249	0.0183	-0.0333	0.0049	-0.0224	0.0008	0.0138	-0.0106	0.0052	0.0298	-0.0157	-0.0120	0.0179	0.0103	-0.0147
	(0.0016)	(0.0132)	(0.0149)	(0.0121)	(0.0144)	(0.0213)	(0.0234)	(0.0277)	(0.0003)	(0.0049)	(0.0102)	(0.0123)	(0.0128)	(0.0203)	(0.0221)	(0.0257)	(0.0242)	(0.0276)
13	-0.0074	-0.0303	-0.0260	-0.0209	0.0165	-0.0424	0.0373	-0.0251	0.0009	0.0073	-0.0038	0.0132	0.0221	-0.0124	0.0064	0.0188	0.0135	0.0062
	(0.0012)	(0.0108)	(0.0108)	(0.0110)	(0.0144)	(0.0185)	(0.0223)	(0.0198)	(0.0002)	(0.0040)	(0.0060)	(0.0079)	(0.0072)	(0.0093)	(0.0096)	(0.0137)	(0.0129)	(0.0157)
14	-0.0047	-0.0414	-0.0172	-0.0303	-0.0342	0.0336	0.0520	0.0105	0.0001	0.0210	-0.0139	0.0122	0.0287	0.0066	0.0141	0.0393	0.0369	0.0197
	(0.0021)	(0.0170)	(0.0163)	(0.0161)	(0.0233)	(0.0291)	(0.0334)	(0.0298)	(0.0004)	(0.0053)	(0.0109)	(0.0117)	(0.0131)	(0.0125)	(0.0139)	(0.0219)	(0.0172)	(0.0177)
15	-0.0076	-0.0464	-0.0348	-0.0355	0.0219	-0.0078	0.0132	-0.0359	0.0001	0.0153	-0.0130	0.0106	0.0414	-0.0259	-0.0014	-0.0018	-0.0191	-0.0124
	(0.0019)	(0.0147)	(0.0125)	(0.0145)	(0.0186)	(0.0195)	(0.0226)	(0.0251)	(0.0003)	(0.0043)	(0.0085)	(0.0112)	(0.0123)	(0.0112)	(0.0117)	(0.0144)	(0.0160)	(0.0205)
16	-0.0074	-0.0195	-0.0120	-0.0130	0.0270	-0.0437	0.0186	-0.0592	0.0007	0.0123	-0.0012	0.0101	0.0157	-0.0191	-0.0033	0.0073	-0.0016	-0.0189
	(0.0017)	(0.0132)	(0.0128)	(0.0120)	(0.0175)	(0.0245)	(0.0309)	(0.0286)	(0.0003)	(0.0040)	(0.0076)	(0.0098)	(0.0115)	(0.0118)	(0.0125)	(0.0176)	(0.0158)	(0.0182)
17	-0.0029	-0.0348	-0.0214	-0.0093	-0.0080	-0.0295	-0.0109	-0.0260	0.0002	0.0239	0.0136	0.0112	0.0133	-0.0066	0.0114	0.0008	0.0038	0.0235
	(0.0009)	(0.0095)	(0.0098)	(0.0099)	(0.0092)	(0.0144)	(0.0159)	(0.0163)	(0.0002)	(0.0036)	(0.0064)	(0.0065)	(0.0073)	(0.0086)	(0.0100)	(0.0111)	(0.0129)	(0.0121)

18	-0.0044	-0.0373	-0.0380	-0.0524	-0.0220	-0.0361	-0.0052	-0.0276	0.0005	0.0287	0.0011	-0.0009	0.0275	0.0003	0.0192	0.0306	0.0340	0.0502
	(0.0015)	(0.0140)	(0.0142)	(0.0151)	(0.0156)	(0.0266)	(0.0312)	(0.0276)	(0.0003)	(0.0053)	(0.0098)	(0.0126)	(0.0109)	(0.0119)	(0.0137)	(0.0148)	(0.0236)	(0.0184)
19	-0.0058	-0.0515	-0.0311	-0.0217	-0.0185	-0.0464	-0.0093	-0.0477	0.0007	0.0291	-0.0018	0.0263	0.0137	-0.0125	0.0128	-0.0165	0.0118	0.0339
	(0.0013)	(0.0108)	(0.0129)	(0.0134)	(0.0109)	(0.0183)	(0.0183)	(0.0190)	(0.0003)	(0.0049)	(0.0092)	(0.0112)	(0.0082)	(0.0123)	(0.0161)	(0.0180)	(0.0179)	(0.0198)
20	-0.0044	-0.0361	-0.0255	-0.0183	-0.0214	-0.0336	-0.0350	-0.0459	0.0006	0.0171	0.0109	0.0243	0.0241	-0.0057	-0.0055	-0.0075	-0.0248	0.0221
	(0.0012)	(0.0116)	(0.0153)	(0.0132)	(0.0109)	(0.0219)	(0.0217)	(0.0224)	(0.0004)	(0.0042)	(0.0086)	(0.0137)	(0.0103)	(0.0111)	(0.0124)	(0.0137)	(0.0179)	(0.0166)
21	-0.0064	-0.0375	-0.0169	-0.0058	-0.0132	-0.0113	-0.0136	-0.0223	0.0002	0.0090	0.0134	0.0038	0.0146	-0.0021	0.0114	0.0257	0.0028	0.0280
	(0.0009)	(0.0086)	(0.0090)	(0.0091)	(0.0083)	(0.0140)	(0.0160)	(0.0146)	(0.0002)	(0.0035)	(0.0053)	(0.0070)	(0.0060)	(0.0075)	(0.0086)	(0.0120)	(0.0114)	(0.0120)
22	-0.0068	-0.0261	-0.0077	-0.0177	-0.0083	-0.0096	0.0032	0.0087	0.0003	0.0142	0.0073	0.0101	0.0164	0.0096	-0.0043	0.0363	0.0330	0.0432
	(0.0020)	(0.0140)	(0.0133)	(0.0133)	(0.0154)	(0.0265)	(0.0333)	(0.0338)	(0.0004)	(0.0059)	(0.0089)	(0.0121)	(0.0105)	(0.0119)	(0.0138)	(0.0145)	(0.0210)	(0.0246)
23	-0.0062	-0.0471	-0.0359	-0.0255	-0.0216	-0.0268	0.0016	-0.0469	0.0006	0.0091	-0.0031	0.0192	0.0070	-0.0001	0.0037	-0.0011	0.0148	0.0405
	(0.0016)	(0.0126)	(0.0143)	(0.0137)	(0.0144)	(0.0197)	(0.0271)	(0.0225)	(0.0003)	(0.0048)	(0.0088)	(0.0110)	(0.0087)	(0.0116)	(0.0125)	(0.0170)	(0.0186)	(0.0234)
24	-0.0072	-0.0414	-0.0235	-0.0176	-0.0346	-0.0185	-0.0371	-0.0421	0.0007	-0.0100	0.0135	0.0060	0.0123	0.0147	0.0097	0.0321	0.0013	0.0543
	(0.0012)	(0.0142)	(0.0162)	(0.0166)	(0.0136)	(0.0252)	(0.0266)	(0.0300)	(0.0003)	(0.0062)	(0.0086)	(0.0106)	(0.0092)	(0.0123)	(0.0128)	(0.0194)	(0.0142)	(0.0262)
25	-0.0042	-0.0218	-0.0132	-0.0118	-0.0078	-0.0088	0.0122	-0.0025	0.0003	0.0095	0.0044	0.0076	0.0193	0.0021	0.0166	0.0231	0.0183	0.0254
	(0.0008)	(0.0083)	(0.0086)	(0.0085)	(0.0085)	(0.0173)	(0.0193)	(0.0203)	(0.0002)	(0.0035)	(0.0048)	(0.0070)	(0.0059)	(0.0077)	(0.0088)	(0.0098)	(0.0110)	(0.0114)
26	-0.0051	-0.0163	0.0036	-0.0119	0.0063	-0.0020	-0.0082	-0.0617	-0.0001	0.0216	-0.0071	0.0131	0.0353	0.0157	0.0261	0.0407	0.0255	0.0651
	(0.0016)	(0.0128)	(0.0146)	(0.0129)	(0.0155)	(0.0291)	(0.0293)	(0.0363)	(0.0004)	(0.0057)	(0.0093)	(0.0101)	(0.0096)	(0.0115)	(0.0127)	(0.0175)	(0.0204)	(0.0225)
27	-0.0075	-0.0388	-0.0233	-0.0247	-0.0125	-0.0312	-0.0136	-0.0588	0.0005	0.0163	-0.0241	0.0141	0.0290	-0.0091	0.0036	0.0162	0.0144	0.0257
	(0.0012)	(0.0122)	(0.0124)	(0.0126)	(0.0146)	(0.0177)	(0.0215)	(0.0170)	(0.0003)	(0.0042)	(0.0073)	(0.0086)	(0.0085)	(0.0122)	(0.0124)	(0.0176)	(0.0171)	(0.0237)
28	-0.0053	-0.0296	-0.0099	-0.0172	-0.0166	-0.0248	-0.0330	-0.0475	0.0002	0.0079	-0.0027	0.0045	0.0264	0.0009	0.0042	0.0135	0.0060	0.0157
	(0.0012)	(0.0106)	(0.0128)	(0.0119)	(0.0100)	(0.0239)	(0.0245)	(0.0299)	(0.0003)	(0.0042)	(0.0081)	(0.0092)	(0.0082)	(0.0126)	(0.0144)	(0.0168)	(0.0165)	(0.0168)
29	-0.0042	-0.0338	-0.0248	-0.0166	-0.0154	-0.0250	-0.0041	-0.0175	0.0003	0.0156	0.0130	0.0136	0.0181	-0.0032	0.0101	0.0145	0.0039	0.0125
	(0.0009)	(0.0078)	(0.0083)	(0.0088)	(0.0088)	(0.0137)	(0.0143)	(0.0150)	(0.0002)	(0.0031)	(0.0047)	(0.0056)	(0.0056)	(0.0069)	(0.0075)	(0.0100)	(0.0102)	(0.0109)
30	-0.0057	-0.0290	-0.0122	-0.0261	-0.0204	0.0180	0.0203	0.0006	0.0000	0.0252	0.0029	0.0097	0.0272	0.0095	0.0142	0.0391	0.0084	0.0389
	(0.0013)	(0.0119)	(0.0122)	(0.0122)	(0.0137)	(0.0308)	(0.0314)	(0.0326)	(0.0003)	(0.0048)	(0.0074)	(0.0091)	(0.0081)	(0.0095)	(0.0107)	(0.0194)	(0.0176)	(0.0170)
31	-0.0058	-0.0356	-0.0218	-0.0174	-0.0161	-0.0157	0.0014	-0.0370	0.0005	0.0183	-0.0057	0.0279	0.0146	-0.0109	0.0072	-0.0068	0.0094	0.0127
	(0.0012)	(0.0105)	(0.0118)	(0.0113)	(0.0139)	(0.0197)	(0.0236)	(0.0198)	(0.0002)	(0.0034)	(0.0067)	(0.0084)	(0.0069)	(0.0092)	(0.0101)	(0.0140)	(0.0138)	(0.0179)
32	-0.0058	-0.0343	-0.0196	-0.0211	-0.0336	-0.0165	-0.0155	-0.0278	0.0006	0.0096	0.0060	0.0196	0.0194	-0.0021	-0.0044	0.0076	-0.0101	0.0130
	(0.0010)	(0.0095)	(0.0120)	(0.0113)	(0.0106)	(0.0209)	(0.0225)	(0.0226)	(0.0002)	(0.0039)	(0.0068)	(0.0082)	(0.0076)	(0.0097)	(0.0107)	(0.0133)	(0.0128)	(0.0157)

[a] Standard deviation based on 200 bootstrap replications is reported in brackets below the corresponding estimated coefficient.

Appendix C3: Estimated parameters of other factors by the unregularized estimation

Group Index	married	us_born	us_m15	us_m5l15	citizenp	mental	rg_ne	rg_mw	rg_sth
1	0.0090	0.0957	0.0864	0.0333	-0.0014	0.0695	0.0020	0.0162	0.0049
	(0.0059)[a]	(0.0231)	(0.0220)	(0.0237)	(0.0152)	(0.0213)	(0.0084)	(0.0088)	(0.0075)
2	-0.0153	0.1458	0.1024	0.0672	0.0493	0.0813	0.0364	0.0331	0.0593
	(0.0114)	(0.0447)	(0.0434)	(0.0391)	(0.0262)	(0.0328)	(0.0181)	(0.0180)	(0.0146)
3	0.0157	0.1078	0.0284	-0.0250	0.0086	0.0657	0.0085	0.0319	0.0410
	(0.0079)	(0.0219)	(0.0216)	(0.0191)	(0.0145)	(0.0227)	(0.0123)	(0.0100)	(0.0103)
4	0.0053	0.1101	0.0801	0.0184	-0.0018	0.0739	0.0078	0.0207	0.0243
	(0.0081)	(0.0248)	(0.0247)	(0.0247)	(0.0160)	(0.0224)	(0.0099)	(0.0095)	(0.0083)
5	-0.0086	0.0921	0.0623	0.0013	0.0078	0.0657	0.0007	0.0107	0.0019
	(0.0062)	(0.0284)	(0.0271)	(0.0275)	(0.0149)	(0.0239)	(0.0080)	(0.0087)	(0.0070)
6	-0.0218	0.1798	0.1064	0.0604	0.0035	-0.0128	0.0336	0.0539	0.0587
	(0.0114)	(0.0264)	(0.0262)	(0.0241)	(0.0208)	(0.0259)	(0.0157)	(0.0154)	(0.0126)
7	-0.0064	0.1496	0.0583	0.0046	-0.0071	0.0444	0.0152	0.0301	0.0294
	(0.0080)	(0.0220)	(0.0222)	(0.0219)	(0.0138)	(0.0343)	(0.0132)	(0.0115)	(0.0107)
8	-0.0073	0.1287	0.0875	0.0159	-0.0083	0.0464	0.0053	0.0240	0.0277
	(0.0086)	(0.0270)	(0.0268)	(0.0234)	(0.0204)	(0.0261)	(0.0115)	(0.0114)	(0.0092)
9	-0.0080	0.0709	0.0381	-0.0133	0.0024	0.0648	-0.0002	0.0108	-0.0019
	(0.0059)	(0.0266)	(0.0245)	(0.0256)	(0.0133)	(0.0233)	(0.0085)	(0.0088)	(0.0066)
10	-0.0292	0.1714	0.1093	0.0922	0.0088	0.0960	0.0327	0.0345	0.0482
	(0.0106)	(0.0311)	(0.0308)	(0.0315)	(0.0296)	(0.0381)	(0.0181)	(0.0163)	(0.0122)
11	-0.0065	0.1331	0.0216	-0.0266	0.0006	0.1011	0.0103	0.0399	0.0483
	(0.0087)	(0.0303)	(0.0293)	(0.0295)	(0.0142)	(0.0390)	(0.0121)	(0.0127)	(0.0093)
12	-0.0205	0.0613	-0.0096	-0.0669	-0.0124	0.0773	0.0027	0.0018	0.0042
	(0.0074)	(0.0569)	(0.0573)	(0.0565)	(0.0167)	(0.0306)	(0.0125)	(0.0115)	(0.0097)
13	-0.0070	0.0834	0.0658	0.0125	0.0115	0.0498	0.0079	0.0146	0.0059
	(0.0060)	(0.0242)	(0.0223)	(0.0230)	(0.0144)	(0.0205)	(0.0089)	(0.0086)	(0.0064)
14	-0.0318	0.1603	0.1090	0.0654	0.0256	0.0417	0.0329	0.0355	0.0561
	(0.0092)	(0.0257)	(0.0249)	(0.0229)	(0.0194)	(0.0246)	(0.0159)	(0.0144)	(0.0123)
15	-0.0091	0.1306	0.0350	-0.0112	0.0120	0.0678	0.0181	0.0295	0.0362
	(0.0070)	(0.0208)	(0.0204)	(0.0196)	(0.0133)	(0.0258)	(0.0103)	(0.0101)	(0.0089)
16	-0.0175	0.0990	0.0559	-0.0074	0.0012	0.0578	0.0064	0.0175	0.0201
	(0.0069)	(0.0295)	(0.0293)	(0.0294)	(0.0171)	(0.0231)	(0.0111)	(0.0107)	(0.0090)
17	0.0344	0.0502	0.0513	0.0223	-0.0015	0.0335	0.0110	0.0162	0.0105
	(0.0056)	(0.0237)	(0.0222)	(0.0227)	(0.0105)	(0.0251)	(0.0077)	(0.0072)	(0.0059)
18	0.0203	0.1024	0.0584	0.0289	-0.0199	0.0557	0.0450	0.0086	0.0488
	(0.0084)	(0.0237)	(0.0201)	(0.0234)	(0.0184)	(0.0243)	(0.0133)	(0.0126)	(0.0099)
19	0.0250	0.0776	0.0110	0.0009	-0.0181	0.0495	0.0241	0.0276	0.0354
	(0.0068)	(0.0173)	(0.0151)	(0.0166)	(0.0126)	(0.0198)	(0.0088)	(0.0099)	(0.0100)
20	0.0412	0.0287	-0.0033	0.0003	0.0351	0.0041	-0.0188	-0.0096	-0.0060
	(0.0078)	(0.0368)	(0.0304)	(0.0293)	(0.0232)	(0.0431)	(0.0108)	(0.0100)	(0.0103)
21	0.0223	0.0426	0.0346	0.0011	0.0222	0.0639	-0.0028	-0.0033	0.0152
	(0.0057)	(0.0198)	(0.0186)	(0.0184)	(0.0132)	(0.0203)	(0.0068)	(0.0065)	(0.0056)
22	0.0157	0.1547	0.0892	0.0528	-0.0126	0.0496	0.0202	0.0134	0.0331
	(0.0078)	(0.0247)	(0.0228)	(0.0211)	(0.0188)	(0.0226)	(0.0139)	(0.0118)	(0.0103)
23	0.0150	0.1193	0.0403	0.0204	-0.0178	0.0484	0.0177	0.0120	0.0225
	(0.0084)	(0.0213)	(0.0201)	(0.0170)	(0.0165)	(0.0243)	(0.0108)	(0.0097)	(0.0089)
24	0.0337	0.0411	0.0002	-0.0282	0.0347	0.0796	-0.0119	-0.0094	0.0015
	(0.0085)	(0.0574)	(0.0510)	(0.0450)	(0.0220)	(0.0364)	(0.0099)	(0.0112)	(0.0095)
25	0.0196	0.0756	0.0545	0.0106	0.0041	0.0115	0.0083	0.0140	0.0121
	(0.0047)	(0.0271)	(0.0267)	(0.0273)	(0.0116)	(0.0201)	(0.0067)	(0.0067)	(0.0056)
26	0.0061	0.0999	0.0372	0.0036	0.0052	0.0227	0.0304	0.0123	0.0295
	(0.0088)	(0.0317)	(0.0318)	(0.0309)	(0.0185)	(0.0386)	(0.0121)	(0.0097)	(0.0094)
27	0.0171	0.1096	0.0140	-0.0227	-0.0134	0.0614	0.0117	0.0122	0.0163
	(0.0067)	(0.0187)	(0.0175)	(0.0171)	(0.0126)	(0.0206)	(0.0091)	(0.0090)	(0.0078)
28	0.0235	0.0631	0.0074	-0.0325	0.0184	0.0361	-0.0087	-0.0047	-0.0037
	(0.0084)	(0.0327)	(0.0306)	(0.0291)	(0.0138)	(0.0195)	(0.0097)	(0.0095)	(0.0090)
29	0.0236	0.0632	0.0503	0.0087	-0.0022	0.0389	0.0112	0.0089	0.0140
	(0.0049)	(0.0194)	(0.0173)	(0.0183)	(0.0116)	(0.0170)	(0.0063)	(0.0054)	(0.0055)
30	0.0090	0.1270	0.0653	0.0378	-0.0103	0.0411	0.0321	0.0146	0.0407
	(0.0068)	(0.0213)	(0.0194)	(0.0206)	(0.0141)	(0.0250)	(0.0107)	(0.0090)	(0.0089)
31	0.0161	0.0941	0.0057	-0.0030	0.0044	0.0525	0.0289	0.0206	0.0270
	(0.0059)	(0.0173)	(0.0164)	(0.0149)	(0.0148)	(0.0184)	(0.0077)	(0.0068)	(0.0076)
32	0.0253	0.0438	0.0051	-0.0250	0.0278	0.0515	-0.0050	-0.0031	0.0009
	(0.0063)	(0.0438)	(0.0402)	(0.0369)	(0.0150)	(0.0257)	(0.0079)	(0.0078)	(0.0076)

[a] Standard deviation based on 200 bootstrap replications is reported in brackets below the corresponding estimated coefficient.

Appendix D1: Full results from the post-selection estimation on relevant lifestyle factors

Group Index	vig_l2	vig_l3	mod_l3	str_l2	str_l3	smk_ed	smk_f	cigsday	alc1yr	alc_life	alc_c1	alc_c2	alc_c3	alc_c4
1	-0.0171	-0.0416	-0.0181	-0.0323	-0.0276	0.0043	0.0239	-0.0001	0.0224	0.0395	0.0035	-0.0319	-0.0552	-0.0596
	(0.0077)[a]	(0.0087)	(0.0069)	(0.0085)	(0.0111)	(0.0218)	(0.0075)	(0.0013)	(0.0141)	(0.0133)	(0.0117)	(0.0120)	(0.0145)	(0.0151)
2	0.0123	0.0000	0.0034	-0.0402	-0.0780	0.0033	0.0241	-0.0013	0.0410	0.0785	-0.0455	-0.0282	-0.0466	-0.0642
	(0.0157)	(0.0169)	(0.0126)	(0.0161)	(0.0187)	(0.0271)	(0.0136)	(0.0018)	(0.0278)	(0.0198)	(0.0224)	(0.0248)	(0.0275)	(0.0302)
3	-0.0169	-0.0199	-0.0111	-0.0188	-0.0149	0.0363	0.0336	-0.0008	0.0421	0.0245	-0.0101	-0.0418	-0.0680	-0.1011
	(0.0107)	(0.0114)	(0.0101)	(0.0109)	(0.0127)	(0.0183)	(0.0113)	(0.0013)	(0.0177)	(0.0166)	(0.0157)	(0.0162)	(0.0189)	(0.0240)
4	-0.0056	-0.0195	-0.0156	-0.0432	-0.0362	0.0137	0.0199	-0.0006	0.0445	0.0426	-0.0133	-0.0443	-0.0717	-0.0773
	(0.0087)	(0.0113)	(0.0078)	(0.0099)	(0.0208)	(0.0182)	(0.0095)	(0.0011)	(0.0153)	(0.0143)	(0.0147)	(0.0136)	(0.0148)	(0.0169)
5	-0.0171	-0.0293	-0.0273	-0.0403	-0.0415	-0.0110	0.0202	0.0001	0.0123	0.0244	0.0054	-0.0323	-0.0652	-0.0743
	(0.0072)	(0.0085)	(0.0063)	(0.0079)	(0.0105)	(0.0173)	(0.0077)	(0.0011)	(0.0152)	(0.0134)	(0.0128)	(0.0125)	(0.0144)	(0.0154)
6	0.0181	-0.0247	-0.0101	-0.0286	-0.0365	0.0050	0.0121	-0.0024	0.0075	0.0367	-0.0086	-0.0163	-0.0409	-0.0279
	(0.0152)	(0.0151)	(0.0124)	(0.0159)	(0.0166)	(0.0258)	(0.0141)	(0.0018)	(0.0235)	(0.0219)	(0.0223)	(0.0221)	(0.0246)	(0.0302)
7	-0.0227	-0.0128	0.0025	-0.0106	-0.0246	0.0387	0.0241	-0.0018	0.0298	0.0061	0.0094	-0.0337	-0.0547	-0.0760
	(0.0104)	(0.0120)	(0.0092)	(0.0095)	(0.0162)	(0.0191)	(0.0108)	(0.0016)	(0.0198)	(0.0182)	(0.0174)	(0.0176)	(0.0205)	(0.0240)
8	-0.0127	-0.0229	-0.0174	-0.0405	-0.0333	0.0125	0.0187	-0.0010	0.0276	0.0155	-0.0070	-0.0461	-0.0807	-0.0757
	(0.0110)	(0.0103)	(0.0098)	(0.0118)	(0.0172)	(0.0202)	(0.0104)	(0.0013)	(0.0238)	(0.0181)	(0.0205)	(0.0230)	(0.0219)	(0.0266)
9	-0.0238	-0.0394	-0.0241	-0.0460	-0.0561	-0.0088	0.0118	-0.0014	0.0187	0.0124	0.0164	-0.0303	-0.0582	-0.0680
	(0.0073)	(0.0083)	(0.0061)	(0.0078)	(0.0102)	(0.0168)	(0.0064)	(0.0010)	(0.0138)	(0.0120)	(0.0124)	(0.0125)	(0.0136)	(0.0151)
10	-0.0250	-0.0209	-0.0040	-0.0322	-0.0881	-0.0077	0.0094	-0.0026	-0.0181	0.0104	-0.0171	-0.0036	-0.0224	0.0124
	(0.0114)	(0.0149)	(0.0111)	(0.0133)	(0.0177)	(0.0263)	(0.0140)	(0.0017)	(0.0202)	(0.0162)	(0.0173)	(0.0174)	(0.0215)	(0.0298)
11	-0.0199	-0.0297	-0.0111	-0.0319	-0.0445	0.0185	0.0136	-0.0022	0.0113	0.0166	0.0010	-0.0264	-0.0560	-0.0487
	(0.0114)	(0.0108)	(0.0099)	(0.0109)	(0.0150)	(0.0261)	(0.0106)	(0.0013)	(0.0229)	(0.0180)	(0.0178)	(0.0213)	(0.0231)	(0.0269)
12	-0.0120	-0.0319	-0.0196	-0.0502	-0.0570	-0.0161	0.0009	-0.0015	-0.0006	0.0114	0.0030	-0.0217	-0.0472	-0.0409
	(0.0097)	(0.0112)	(0.0080)	(0.0098)	(0.0168)	(0.0198)	(0.0103)	(0.0013)	(0.0200)	(0.0172)	(0.0181)	(0.0189)	(0.0194)	(0.0256)
13	-0.0130	-0.0260	-0.0293	-0.0379	-0.0522	-0.0247	0.0170	0.0002	0.0135	0.0381	0.0143	-0.0281	-0.0637	-0.0732
	(0.0069)	(0.0082)	(0.0062)	(0.0072)	(0.0095)	(0.0165)	(0.0072)	(0.0010)	(0.0142)	(0.0123)	(0.0122)	(0.0127)	(0.0147)	(0.0154)
14	-0.0046	-0.0204	-0.0004	-0.0322	-0.0728	0.0044	0.0241	-0.0025	0.0027	0.0392	-0.0333	-0.0196	-0.0391	-0.0362
	(0.0114)	(0.0132)	(0.0101)	(0.0127)	(0.0134)	(0.0227)	(0.0122)	(0.0015)	(0.0209)	(0.0158)	(0.0190)	(0.0192)	(0.0234)	(0.0249)
15	-0.0176	-0.0158	-0.0086	-0.0232	-0.0344	0.0218	0.0278	-0.0010	0.0170	0.0038	0.0172	-0.0229	-0.0527	-0.0708
	(0.0087)	(0.0100)	(0.0077)	(0.0083)	(0.0134)	(0.0174)	(0.0094)	(0.0011)	(0.0178)	(0.0215)	(0.0174)	(0.0148)	(0.0178)	(0.0197)
16	-0.0030	-0.0156	-0.0196	-0.0362	-0.0516	-0.0238	0.0152	0.0000	0.0048	0.0334	0.0023	-0.0294	-0.0622	-0.0615
	(0.0088)	(0.0109)	(0.0079)	(0.0109)	(0.0135)	(0.0200)	(0.0094)	(0.0011)	(0.0164)	(0.0164)	(0.0165)	(0.0153)	(0.0163)	(0.0187)
17	-0.0198	-0.0387	-0.0113	-0.0112	-0.0053	-0.0393	0.0258	0.0002	0.0223	0.0335	0.0111	-0.0096	-0.0209	-0.0313
	(0.0068)	(0.0074)	(0.0059)	(0.0063)	(0.0084)	(0.0152)	(0.0058)	(0.0008)	(0.0143)	(0.0111)	(0.0110)	(0.0124)	(0.0120)	(0.0147)
18	-0.0283	-0.0406	-0.0153	-0.0099	0.0193	-0.0426	0.0229	-0.0014	0.0310	0.0581	-0.0314	-0.0189	-0.0179	-0.0189

	(0.0111)	(0.0132)	(0.0102)	(0.0119)	(0.0158)	(0.0181)	(0.0094)	(0.0013)	(0.0214)	(0.0156)	(0.0168)	(0.0194)	(0.0184)	(0.0210)
19	-0.0101	-0.0162	-0.0141	-0.0072	0.0000	-0.0321	0.0155	0.0013	0.0471	0.0138	-0.0162	-0.0431	-0.0397	-0.0784
	(0.0081)	(0.0118)	(0.0096)	(0.0097)	(0.0133)	(0.0157)	(0.0087)	(0.0012)	(0.0152)	(0.0137)	(0.0135)	(0.0125)	(0.0144)	(0.0248)
20	-0.0179	-0.0173	-0.0073	-0.0193	-0.0127	0.0014	0.0119	-0.0025	0.0600	0.0490	-0.0201	-0.0457	-0.0576	-0.0666
	(0.0089)	(0.0134)	(0.0089)	(0.0100)	(0.0138)	(0.0225)	(0.0109)	(0.0009)	(0.0212)	(0.0173)	(0.0216)	(0.0175)	(0.0180)	(0.0212)
21	-0.0179	-0.0359	-0.0118	-0.0159	-0.0122	-0.0207	0.0221	-0.0006	0.0269	0.0171	0.0119	-0.0144	-0.0326	-0.0493
	(0.0060)	(0.0066)	(0.0056)	(0.0058)	(0.0069)	(0.0114)	(0.0064)	(0.0007)	(0.0136)	(0.0106)	(0.0107)	(0.0119)	(0.0118)	(0.0136)
22	-0.0129	-0.0305	-0.0100	-0.0123	0.0173	-0.0216	0.0141	-0.0012	0.0061	0.0540	-0.0177	-0.0105	-0.0150	0.0046
	(0.0104)	(0.0122)	(0.0090)	(0.0109)	(0.0142)	(0.0212)	(0.0102)	(0.0013)	(0.0185)	(0.0180)	(0.0169)	(0.0171)	(0.0173)	(0.0245)
23	-0.0091	-0.0120	-0.0132	-0.0103	-0.0087	-0.0281	0.0069	0.0006	0.0503	0.0339	0.0027	-0.0233	-0.0389	-0.0546
	(0.0092)	(0.0126)	(0.0089)	(0.0100)	(0.0117)	(0.0199)	(0.0093)	(0.0012)	(0.0191)	(0.0160)	(0.0165)	(0.0179)	(0.0187)	(0.0190)
24	-0.0184	-0.0180	-0.0068	-0.0355	-0.0189	0.0112	0.0111	-0.0029	0.0554	0.0474	-0.0333	-0.0320	-0.0580	-0.0461
	(0.0105)	(0.0131)	(0.0102)	(0.0101)	(0.0133)	(0.0185)	(0.0121)	(0.0007)	(0.0201)	(0.0202)	(0.0192)	(0.0186)	(0.0190)	(0.0276)
25	-0.0160	-0.0288	-0.0162	-0.0261	-0.0226	-0.0454	0.0096	-0.0005	0.0016	0.0133	0.0216	0.0042	-0.0159	-0.0336
	(0.0062)	(0.0064)	(0.0056)	(0.0061)	(0.0076)	(0.0115)	(0.0056)	(0.0006)	(0.0111)	(0.0109)	(0.0110)	(0.0096)	(0.0100)	(0.0129)
26	-0.0420	-0.0341	-0.0038	-0.0128	-0.0012	-0.0361	0.0161	-0.0029	0.0206	0.0420	-0.0046	-0.0064	-0.0073	-0.0181
	(0.0104)	(0.0120)	(0.0106)	(0.0108)	(0.0146)	(0.0208)	(0.0110)	(0.0011)	(0.0190)	(0.0158)	(0.0174)	(0.0172)	(0.0170)	(0.0290)
27	-0.0165	-0.0180	-0.0116	-0.0049	-0.0145	-0.0254	0.0110	-0.0013	0.0334	0.0124	0.0054	-0.0178	-0.0306	-0.0525
	(0.0080)	(0.0103)	(0.0077)	(0.0089)	(0.0106)	(0.0154)	(0.0088)	(0.0008)	(0.0159)	(0.0155)	(0.0152)	(0.0139)	(0.0149)	(0.0163)
28	-0.0155	-0.0256	-0.0118	-0.0252	-0.0134	-0.0271	0.0109	-0.0023	0.0134	0.0260	-0.0054	-0.0156	-0.0380	-0.0378
	(0.0085)	(0.0079)	(0.0072)	(0.0076)	(0.0105)	(0.0116)	(0.0082)	(0.0005)	(0.0178)	(0.0194)	(0.0162)	(0.0134)	(0.0137)	(0.0176)
29	-0.0226	-0.0364	-0.0152	-0.0167	-0.0115	-0.0253	0.0136	-0.0008	0.0228	0.0171	0.0328	-0.0105	-0.0215	-0.0360
	(0.0055)	(0.0061)	(0.0052)	(0.0051)	(0.0070)	(0.0091)	(0.0061)	(0.0006)	(0.0111)	(0.0092)	(0.0107)	(0.0105)	(0.0102)	(0.0124)
30	-0.0357	-0.0466	-0.0094	-0.0120	0.0142	-0.0266	0.0167	-0.0021	0.0231	0.0495	-0.0141	-0.0144	-0.0202	-0.0272
	(0.0089)	(0.0103)	(0.0077)	(0.0090)	(0.0115)	(0.0153)	(0.0107)	(0.0009)	(0.0176)	(0.0134)	(0.0153)	(0.0174)	(0.0171)	(0.0255)
31	-0.0239	-0.0262	-0.0120	-0.0051	-0.0111	-0.0192	0.0082	-0.0004	0.0394	0.0208	0.0125	-0.0212	-0.0267	-0.0494
	(0.0071)	(0.0100)	(0.0075)	(0.0083)	(0.0104)	(0.0147)	(0.0085)	(0.0008)	(0.0155)	(0.0126)	(0.0127)	(0.0142)	(0.0153)	(0.0169)
32	-0.0194	-0.0255	-0.0147	-0.0241	-0.0060	-0.0171	0.0132	-0.0019	0.0354	0.0397	-0.0045	-0.0249	-0.0467	-0.0439
	(0.0081)	(0.0093)	(0.0081)	(0.0067)	(0.0105)	(0.0217)	(0.0088)	(0.0009)	(0.0142)	(0.0145)	(0.0130)	(0.0113)	(0.0122)	(0.0161)

[a] Standard deviation based on 200 bootstrap replications for post-selection estimation is reported in brackets below the corresponding estimated coefficient.

Appendix D2: Full results from the post-selection estimation on relevant socio-economic factors and other factors

Group Index	educ1	occup1	occup2	lnincome	hp	married	us_born	us_m15	mental	rg_sth
1	-0.0098	-0.0087	-0.0213	0.0218	0.0092	0.0115	0.0611	0.0534	0.0692	-0.0010
	(0.0011)[a]	(0.0067)	(0.0085)	(0.0036)	(0.0092)	(0.0057)	(0.0122)	(0.0138)	(0.0181)	(0.0064)
2	-0.0084	-0.0081	0.0145	0.0301	-0.0037	-0.0179	0.1124	0.0564	0.0775	0.0360
	(0.0026)	(0.0148)	(0.0147)	(0.0071)	(0.0177)	(0.0098)	(0.0177)	(0.0207)	(0.0356)	(0.0107)
3	-0.0115	-0.0275	-0.0242	0.0274	0.0177	0.0147	0.1278	0.0456	0.0681	0.0321
	(0.0019)	(0.0123)	(0.0126)	(0.0048)	(0.0094)	(0.0089)	(0.0122)	(0.0126)	(0.0255)	(0.0090)
4	-0.0088	-0.0162	-0.0058	0.0261	-0.0071	0.0052	0.0869	0.0578	0.0748	0.0161
	(0.0019)	(0.0083)	(0.0126)	(0.0036)	(0.0170)	(0.0084)	(0.0130)	(0.0149)	(0.0197)	(0.0071)
5	-0.0107	-0.0247	-0.0301	0.0075	0.0103	-0.0067	0.0949	0.0594	0.0667	-0.0017
	(0.0011)	(0.0068)	(0.0087)	(0.0037)	(0.0072)	(0.0051)	(0.0117)	(0.0131)	(0.0207)	(0.0059)
6	-0.0058	-0.0326	0.0161	0.0115	0.0223	-0.0260	0.1294	0.0514	-0.0091	0.0301
	(0.0023)	(0.0141)	(0.0140)	(0.0061)	(0.0158)	(0.0105)	(0.0167)	(0.0198)	(0.0259)	(0.0099)
7	-0.0095	-0.0341	-0.0274	0.0092	0.0288	-0.0108	0.1354	0.0429	0.0428	0.0187
	(0.0022)	(0.0125)	(0.0115)	(0.0045)	(0.0100)	(0.0082)	(0.0113)	(0.0128)	(0.0341)	(0.0084)
8	-0.0089	-0.0304	-0.0311	0.0109	0.0009	-0.0081	0.1085	0.0661	0.0469	0.0185
	(0.0022)	(0.0105)	(0.0123)	(0.0048)	(0.0142)	(0.0085)	(0.0154)	(0.0191)	(0.0232)	(0.0089)
9	-0.0081	-0.0077	-0.0093	0.0075	0.0197	-0.0071	0.0813	0.0458	0.0662	-0.0054
	(0.0010)	(0.0068)	(0.0084)	(0.0032)	(0.0087)	(0.0055)	(0.0118)	(0.0125)	(0.0233)	(0.0059)
10	-0.0064	-0.0061	0.0095	0.0199	0.0201	-0.0328	0.0915	0.0267	0.0939	0.0266
	(0.0019)	(0.0127)	(0.0115)	(0.0051)	(0.0174)	(0.0092)	(0.0217)	(0.0233)	(0.0452)	(0.0093)
11	-0.0089	-0.0028	0.0065	0.0137	0.0323	-0.0131	0.1469	0.0308	0.0961	0.0358
	(0.0016)	(0.0119)	(0.0130)	(0.0043)	(0.0121)	(0.0088)	(0.0124)	(0.0154)	(0.0444)	(0.0086)
12	-0.0093	-0.0069	0.0289	0.0177	0.0253	-0.0233	0.1008	0.0339	0.0811	0.0037
	(0.0016)	(0.0096)	(0.0137)	(0.0036)	(0.0112)	(0.0075)	(0.0157)	(0.0183)	(0.0290)	(0.0074)
13	-0.0069	-0.0151	-0.0176	0.0117	0.0200	-0.0038	0.0738	0.0498	0.0557	0.0006
	(0.0011)	(0.0069)	(0.0072)	(0.0032)	(0.0074)	(0.0056)	(0.0111)	(0.0119)	(0.0176)	(0.0060)
14	-0.0055	-0.0188	0.0076	0.0230	0.0268	-0.0344	0.1121	0.0556	0.0440	0.0337
	(0.0021)	(0.0118)	(0.0111)	(0.0048)	(0.0126)	(0.0086)	(0.0154)	(0.0182)	(0.0289)	(0.0094)
15	-0.0073	-0.0201	-0.0125	0.0169	0.0369	-0.0095	0.1438	0.0410	0.0673	0.0249
	(0.0020)	(0.0104)	(0.0116)	(0.0037)	(0.0122)	(0.0085)	(0.0125)	(0.0140)	(0.0296)	(0.0072)
16	-0.0076	-0.0119	-0.0062	0.0161	0.0114	-0.0184	0.1017	0.0556	0.0616	0.0137
	(0.0017)	(0.0091)	(0.0103)	(0.0046)	(0.0120)	(0.0077)	(0.0132)	(0.0146)	(0.0206)	(0.0073)
17	-0.0035	-0.0299	-0.0175	0.0266	0.0137	0.0402	0.0303	0.0310	0.0321	0.0019
	(0.0009)	(0.0067)	(0.0081)	(0.0028)	(0.0063)	(0.0056)	(0.0093)	(0.0103)	(0.0246)	(0.0058)
18	-0.0053	-0.0108	-0.0124	0.0355	0.0330	0.0281	0.0655	0.0290	0.0547	0.0369
	(0.0014)	(0.0117)	(0.0122)	(0.0044)	(0.0104)	(0.0089)	(0.0145)	(0.0163)	(0.0189)	(0.0086)
19	-0.0071	-0.0412	-0.0178	0.0339	0.0101	0.0282	0.0603	0.0026	0.0519	0.0234
	(0.0013)	(0.0086)	(0.0119)	(0.0040)	(0.0081)	(0.0067)	(0.0099)	(0.0118)	(0.0190)	(0.0088)
20	-0.0061	-0.0256	-0.0079	0.0241	0.0184	0.0441	0.0506	0.0081	0.0119	0.0020
	(0.0012)	(0.0080)	(0.0144)	(0.0033)	(0.0091)	(0.0094)	(0.0167)	(0.0180)	(0.0333)	(0.0087)
21	-0.0060	-0.0316	-0.0115	0.0132	0.0179	0.0291	0.0627	0.0449	0.0652	0.0175
	(0.0010)	(0.0058)	(0.0073)	(0.0031)	(0.0051)	(0.0050)	(0.0087)	(0.0103)	(0.0180)	(0.0051)
22	-0.0064	-0.0167	-0.0003	0.0170	0.0173	0.0219	0.1037	0.0396	0.0561	0.0232
	(0.0017)	(0.0101)	(0.0120)	(0.0060)	(0.0095)	(0.0089)	(0.0119)	(0.0145)	(0.0253)	(0.0085)
23	-0.0073	-0.0317	-0.0174	0.0138	0.0040	0.0168	0.0907	0.0154	0.0411	0.0151
	(0.0013)	(0.0090)	(0.0114)	(0.0042)	(0.0080)	(0.0071)	(0.0111)	(0.0130)	(0.0219)	(0.0078)
24	-0.0070	-0.0220	-0.0090	0.0015	0.0146	0.0402	0.0944	0.0438	0.0734	0.0086
	(0.0012)	(0.0097)	(0.0126)	(0.0051)	(0.0093)	(0.0097)	(0.0147)	(0.0154)	(0.0324)	(0.0082)
25	-0.0044	-0.0147	-0.0075	0.0110	0.0214	0.0244	0.0715	0.0473	0.0131	0.0053
	(0.0008)	(0.0055)	(0.0072)	(0.0030)	(0.0058)	(0.0051)	(0.0091)	(0.0098)	(0.0208)	(0.0049)
26	-0.0051	-0.0162	0.0069	0.0242	0.0365	0.0107	0.0940	0.0303	0.0185	0.0172
	(0.0015)	(0.0105)	(0.0125)	(0.0053)	(0.0093)	(0.0080)	(0.0125)	(0.0136)	(0.0329)	(0.0086)
27	-0.0076	-0.0264	-0.0093	0.0171	0.0263	0.0140	0.1076	0.0140	0.0595	0.0110
	(0.0012)	(0.0084)	(0.0092)	(0.0040)	(0.0082)	(0.0068)	(0.0102)	(0.0115)	(0.0262)	(0.0062)
28	-0.0057	-0.0197	0.0014	0.0103	0.0259	0.0246	0.0994	0.0404	0.0275	0.0001
	(0.0012)	(0.0090)	(0.0094)	(0.0038)	(0.0085)	(0.0078)	(0.0111)	(0.0120)	(0.0181)	(0.0065)
29	-0.0044	-0.0212	-0.0158	0.0170	0.0191	0.0296	0.0572	0.0436	0.0394	0.0082
	(0.0009)	(0.0055)	(0.0066)	(0.0023)	(0.0052)	(0.0045)	(0.0087)	(0.0101)	(0.0162)	(0.0048)
30	-0.0059	-0.0149	0.0010	0.0275	0.0278	0.0128	0.0894	0.0297	0.0394	0.0269
	(0.0012)	(0.0093)	(0.0109)	(0.0045)	(0.0079)	(0.0075)	(0.0105)	(0.0122)	(0.0232)	(0.0070)
31	-0.0059	-0.0264	-0.0080	0.0224	0.0099	0.0165	0.0977	0.0066	0.0540	0.0158
	(0.0014)	(0.0068)	(0.0081)	(0.0033)	(0.0068)	(0.0059)	(0.0095)	(0.0117)	(0.0191)	(0.0070)
32	-0.0057	-0.0163	-0.0021	0.0133	0.0164	0.0264	0.0825	0.0371	0.0471	0.0051
	(0.0009)	(0.0073)	(0.0094)	(0.0031)	(0.0077)	(0.0071)	(0.0118)	(0.0126)	(0.0216)	(0.0062)

[a] Standard deviation based on 200 bootstrap replications for post-selection estimation is reported in brackets below the corresponding estimated coefficient.

Appendix E: Mathematical Proofs

This document contains proofs for preliminary Lemma and Theorems. Throughout this note, $\text{diag}(V)$ denotes a square diagonal matrix with the elements of the vector V on the main diagonal; $\to_P$ denotes converging in probability; $\to_D$ denotes converging in distribution; $\|\cdot\|$ denotes the Euclidean norm; let $\Pr(z) = \Pr(Z_i = z)$ for notational simplicity when no misunderstanding can arise; i_p denotes a $p \times 1$ one vector; and I_p denotes a $p \times p$ identity matrix.

For notational simplicity, partition $\hat{\Theta}$ as $\hat{\Theta} = (\hat{\bar{\Theta}}', \hat{\tilde{\Theta}}')'$ conformably with $Z_i = (\bar{Z}_i', \tilde{Z}_i')'$, where $\hat{\bar{\Theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_{\bar{r}})'$ and $\tilde{\Theta} = (\hat{\theta}_{\bar{r}+1}, \ldots, \hat{\theta}_r)'$. In addition, for $j = 1, \ldots, m$, let $\bar{z}^j = (z_1^j, \ldots, z_{\bar{r}}^j)'$ and $\tilde{z}^j = (z_{\bar{r}+1}^j, \ldots, z_r^j)'$. In the following proofs, we will repeatly use these notations.

Lemma E.1. *Let $\hat{\Theta} = (\hat{\bar{\Theta}}', \hat{\tilde{\Theta}}')'$ be the one obtained from Lemma 2.1. Under Assumptions 1-3, for $\forall z \in \mathcal{D}$*

1. $\frac{1}{N}\sum_{i=1}^N X_i X_i' L(Z_i, z, \hat{\Theta}) = \Pr(\bar{z}) E[X_i X_i' | \bar{z}] \Delta_1(\tilde{z}) + O_P\left(\frac{1}{\sqrt{N}}\right)$, *where $\Delta_1(\tilde{z})$ is denoted in (E.2);*
2. $\frac{1}{N}\sum_{i=1}^N X_i X_i' \beta(\bar{Z}_i) L(Z_i, z, \hat{\Theta}) = \Pr(\bar{z}) E[X_i X_i' \beta(\bar{z}) | \bar{z}] \Delta_1(\tilde{z}) + O_P\left(\frac{1}{\sqrt{N}}\right)$, *where $\Delta_1(\tilde{z})$ is denoted in (E.2);*
3. $\frac{1}{N}\sum_{i=1}^N X_i \varepsilon_i L(Z_i, z, \hat{\Theta}) = O_P\left(\frac{1}{\sqrt{N}}\right)$;
4. $\frac{1}{N}\sum_{i=1}^N X_i' \beta_0(Z_i) \varepsilon_i L(Z_i, z, \hat{\Theta}) = O_P\left(\frac{1}{\sqrt{N}}\right)$;
5. $\frac{1}{N}\sum_{i=1}^N \varepsilon_i^2 L(Z_i, z, \hat{\Theta}) = \sigma_\varepsilon^2(\bar{z}) \Pr(\bar{z}) \Delta_1(\tilde{z}) + O_P\left(\frac{1}{\sqrt{N}}\right)$, *where $\Delta_1(\tilde{z})$ is denoted in (E.2).*

Proof of Lemma E.1:

1). By the definition of the kernel function used in this paper, we can write for $s = 1, \ldots, r$

$$l(Z_{i,s}, z_s, \hat{\theta}_s) = 1(Z_{i,s} = z_s) + \hat{\theta}_s 1(Z_{i,s} \neq z_s).$$

Based on Lemma 2.1, we can simplify the product kernel as

$$L(Z_i, z, \hat{\Theta}) = \left(1(\bar{Z}_i = \bar{z}) + \sum_{s=1}^{\bar{r}} \hat{\theta}_s 1_{s, \bar{Z}_i = \bar{z}} + O_P(\|\hat{\bar{\Theta}}\|^2)\right) L(\tilde{Z}_i, \tilde{z}, \hat{\tilde{\Theta}}),$$

where $1_{s, \bar{Z}_i = \bar{z}} = 1(Z_{i,s} \neq z_s) \prod_{n=1, n\neq s}^{\bar{r}} 1(Z_{i,n} = z_n)$. Therefore,

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^N X_i X_i' L(Z_i, z, \hat{\Theta}) &= \frac{1}{N}\sum_{i=1}^N X_i X_i' \left(1(\bar{Z}_i = \bar{z}) + \sum_{s=1}^{\bar{r}} \hat{\theta}_s 1_{s, \bar{Z}_i = \bar{z}} + O_P(\|\hat{\bar{\Theta}}\|^2)\right) \tilde{L}(\tilde{Z}_i, \tilde{z}, \hat{\tilde{\Theta}}) \\ &\equiv A_1 + A_2 + A_3, \qquad \text{(E.1)}\end{aligned}$$

where

$$A_1 = \frac{1}{N}\sum_{i=1}^N X_i X_i' 1(\bar{Z}_i = \bar{z}) \tilde{L}(\tilde{Z}_i, \tilde{z}, \hat{\tilde{\Theta}}), \quad A_2 = \frac{1}{N}\sum_{i=1}^N X_i X_i' \sum_{s=1}^{\bar{r}} \hat{\theta}_s 1_{s, \bar{Z}_i = \bar{z}} \tilde{L}(\tilde{Z}_i, \tilde{z}, \hat{\tilde{\Theta}}),$$

$$A_3 = O_P(\|\hat{\hat{\Theta}}\|^2)\frac{1}{N}\sum_{i=1}^{N} X_i X_i' \tilde{L}(\tilde{Z}_i, \tilde{z}, \hat{\hat{\Theta}}).$$

Notice that we can expand the product form of $\tilde{L}(\tilde{Z}_i, \tilde{z}, \hat{\hat{\Theta}})$ as a summation form:

$$\begin{aligned}
\tilde{L}(\tilde{Z}_i, \tilde{z}, \hat{\hat{\Theta}}) &= \prod_{s=\bar{r}+1}^{r} \left(1(Z_{i,s} = z_s) + \hat{\theta}_s 1(Z_{i,s} \neq z_s)\right) \\
&= \prod_{s=\bar{r}+1}^{r} 1(Z_{i,s} = z_s) + \cdots + \prod_{s=\bar{r}+1}^{r} \hat{\theta}_s 1(Z_{i,s} \neq z_s).
\end{aligned}$$

Then, for simplicity, we denote

$$\Delta_1(\tilde{z}) = E\left[\prod_{s=\bar{r}+1}^{r} 1(Z_{i,s} = z_s)\right] + \cdots + \prod_{s=\bar{r}+1}^{r} \hat{\theta}_s E\left[\prod_{s=\bar{r}+1}^{r} 1(Z_{i,s} \neq z_s)\right] \tag{E.2}$$

as the expectation of $\tilde{L}(\tilde{Z}_i, \tilde{z}, \hat{\hat{\Theta}})$ with respect to $\tilde{Z}_i$. In connection with Assumption 1, it is easy to know that

$$A_1 = \Pr(\bar{z}) E[X_i X_i' | \bar{z}] \Delta_1(\tilde{z}) + O_P\left(\frac{1}{\sqrt{N}}\right),$$

where $\Delta_1(\tilde{z})$ is denoted in (E.2).

For A_2,

$$\begin{aligned}
\|A_2\| &\le \sum_{s=1}^{\bar{r}} \left\| \theta_s \frac{1}{N}\sum_{i=1}^{N} X_i X_i' 1_{s,\bar{Z}_i=\bar{z}} \tilde{L}(\tilde{Z}_i, \tilde{z}, \hat{\hat{\Theta}}) \right\| \\
&\le \sum_{s=1}^{\bar{r}} |\hat{\theta}_s| \left\| \frac{1}{N}\sum_{i=1}^{N} X_i X_i' 1_{s,\bar{Z}_i=\bar{z}} \tilde{L}(\tilde{Z}_i, \tilde{z}, \hat{\hat{\Theta}}) \right\| \\
&\le O_P(\|\hat{\hat{\Theta}}\|) \sum_{s=1}^{\bar{r}} \left\| \frac{1}{N}\sum_{i=1}^{N} X_i X_i' 1_{s,\bar{Z}_i=\bar{z}} \tilde{L}(\tilde{Z}_i, \tilde{z}, \hat{\hat{\Theta}}) \right\| = O_P(\|\hat{\hat{\Theta}}\|).
\end{aligned}$$

Similar to A_2, we can show that $A_3 = O_P(\|\hat{\hat{\Theta}}\|^2)$. Based on Lemma 2.1 and the analysis for A_1, A_2 and A_3, the proof is complete. ■

2)-5) The results follow from the procedure similar to 1) of this lemma. ■

Proof of Theorem 2.1:

1). Let $\alpha_N = \frac{1}{\sqrt{N}}$ and U be an $m \times p$ matrix. We want to show that for any given $\epsilon > 0$, there exists a large constant C such that

$$\liminf_{N} \Pr\left\{ \inf_{\|U\|=C} Q_\gamma\left(B_0 + \alpha_N U\right) > Q_\gamma(B_0) \right\} = 1 - \epsilon. \tag{E.3}$$

This implies with probability at least $1-\epsilon$ that there exists a local minimum in the ball $\{B_0 + \alpha_N U : \|U\| \le C\}$. Hence, there exists a local minimizer such that $\|\hat{B} - B_0\| = O_P(\alpha_N)$. The above argument is in line with the same spirit of the proofs for Theorem 1 of Fan and Li (2001) and Lemma A.1 of Wang and Xia (2009).

For notational simplicity, let U_j be the transpose of the jth row of the matrix U with $j = 1, \ldots, m$ and V_s be the sth column of the matrix U with $s = 1, \ldots, p$; and denote

$$\begin{aligned} e_j &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i \left(X_i' \beta_0(\bar{Z}_i) - X_i' \beta_0(\bar{z}^j) + \varepsilon_i \right) L(Z_i, z^j, \hat{\Theta}) \\ &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i X_i' \left(\beta_0(\bar{Z}_i) - \beta_0(\bar{z}^j) \right) L(Z_i, z^j, \hat{\Theta}) + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i \varepsilon_i L(Z_i, z^j, \hat{\Theta}). \end{aligned}$$

By result (3) of Lemma E.1, it is easy to know that $\frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i \varepsilon_i L(Z_i, z^j, \hat{\Theta}) = O_P(1)$ uniformly in j. We now focus on the next term

$$\begin{aligned} & \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i X_i' \left(\beta_0(\bar{Z}_i) - \beta_0(\bar{z}^j) \right) L(Z_i, z^j, \hat{\Theta}) \right\| \\ = & \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i X_i' \left(\beta_0(\bar{Z}_i) - \beta_0(\bar{z}^j) \right) \left(1(\bar{Z}_i = \bar{z}^j) + \sum_{s=1}^{\bar{r}} \hat{\theta}_s 1_{s, \bar{Z}_i = \bar{z}^j} + O_P(\|\hat{\tilde{\Theta}}\|^2) \right) \tilde{L}(\tilde{Z}_i, \tilde{z}, \hat{\tilde{\Theta}}) \right\| \\ \leq & \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i X_i' \left(\beta_0(\bar{Z}_i) - \beta_0(\bar{z}^j) \right) \sum_{s=1}^{\bar{r}} \hat{\theta}_s 1_{s, \bar{Z}_i = \bar{z}^j} \tilde{L}(\tilde{Z}_i, \tilde{z}, \hat{\tilde{\Theta}}) \right\| \\ & + \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i X_i' \left(\beta_0(\bar{Z}_i) - \beta_0(\bar{z}^j) \right) O_P(\|\hat{\tilde{\Theta}}\|^2) \tilde{L}(\tilde{Z}_i, \tilde{z}, \hat{\tilde{\Theta}}) \right\| = O_P(1), \end{aligned} \quad \text{(E.4)}$$

where the last equality follows from Lemma 2.1, Assumption 1, and results (1) and (2) of Lemma E.1. Therefore, we know that $e_j = O_P(1)$ uniformly in j due to the fact that $\mathcal{D}$ is compact.

Then we write

$$\begin{aligned} & Q_\gamma(B_0 + \alpha_N U) - Q_\gamma(B_0) \\ = & \sum_{j=1}^{m} \sum_{i=1}^{N} \left(X_i' \beta_0(\bar{Z}_i) + \varepsilon_i - X_i' \beta_0(\bar{z}^j) - \alpha_N X_i' U_j \right)^2 L(Z_i, z^j, \hat{\Theta}) \\ & + \sum_{s=1}^{p^*} \gamma_s \|b_{0s} + \alpha_N V_s\| + \sum_{s=p^*+1}^{p} \gamma_s \|\alpha_N V_s\| \\ & - \sum_{j=1}^{m} \sum_{i=1}^{N} \left(X_i' \beta_0(\bar{Z}_i) + \varepsilon_i - X_i' \beta_0(\bar{z}^j) \right)^2 L(Z_i, z^j, \hat{\Theta}) - \sum_{s=1}^{p^*} \gamma_s \|b_{0s}\| \\ = & \sum_{j=1}^{m} \sum_{i=1}^{N} \left(\alpha_N X_i' U_j \right)^2 L(Z_i, z^j, \hat{\Theta}) + \sum_{s=p^*+1}^{p} \gamma_s \|\alpha_N V_s\| + \sum_{s=1}^{p^*} \gamma_s \left(\|b_{0s} + \alpha_N V_s\| - \|b_{0s}\| \right) \\ & - 2 \sum_{j=1}^{m} \sum_{i=1}^{N} \alpha_N U_j' X_i \left(X_i' \beta_0(\bar{Z}_i) - X_i' \beta_0(\bar{z}^j) + \varepsilon_i \right) L(Z_i, z^j, \hat{\Theta}) \\ \geq & \sum_{j=1}^{m} \sum_{i=1}^{N} \alpha_N^2 U_j' X_i X_i' U_j L(Z_i, z^j, \hat{\Theta}) + \sum_{s=1}^{p^*} \gamma_s \left(\|b_{0s} + \alpha_N V_s\| - \|b_{0s}\| \right) \\ & - 2 \sum_{j=1}^{m} \sum_{i=1}^{N} \alpha_N U_j' X_i \left(X_i' \beta_0(\bar{Z}_i) - X_i' \beta_0(\bar{z}^j) + \varepsilon_i \right) L(Z_i, z^j, \hat{\Theta}) \end{aligned}$$

$$
\begin{aligned}
&\geq \sum_{j=1}^{m} \|U_j\|^2 \frac{\Lambda_{min}(z^j)}{2} - 2\sum_{j=1}^{m} U_j' e_j + \sum_{s=1}^{p^*} \gamma_s \left(\|b_{0s} + \alpha_N V_s\| - \|b_{0s}\|\right) \\
&\geq \sum_{j=1}^{m} \|U_j\|^2 \frac{\Lambda_{min}(z^j)}{2} - 2\sum_{j=1}^{m} U_j' e_j - O(1) \sum_{s=1}^{p^*} \gamma_s \frac{1}{\sqrt{N}} \|V_s\|,
\end{aligned}
$$

where $\Lambda_{min}(z^j)$ denotes the minimum eigenvalue of $\Pr(\bar{z}^j)E[X_iX_i'|\bar{z}^j]\Delta_1(\bar{z}^j)$; the second inequality follows from (1) of Lemma E.1; the third inequality follows from the Mean Value Theorem. Notice that $\|U\| = C$, so we further write that

$$
\begin{aligned}
& Q_\gamma(B_0 + \alpha_N U) - Q_\gamma(B_0) \\
\geq & \sum_{j=1}^{m} \|U_j\|^2 \frac{\Lambda_{min}(z^j)}{2} - 2\sum_{j=1}^{m} U_j' e_j - O(1) \sum_{s=1}^{p^*} \gamma_s \frac{1}{\sqrt{N}} \|V_s\| \\
\geq & \sum_{j=1}^{m} \|U_j\|^2 \frac{\Lambda_{min}(z^j)}{2} - 2 \left(\sum_{j=1}^{m} \|U_j\|^2 \sum_{j=1}^{m} \|e_j\|^2 \right)^{1/2} - O(1) \sum_{s=1}^{p^*} \gamma_s \frac{1}{\sqrt{N}} \|V_s\| \\
\geq & C^2 \min_j \frac{\Lambda_{min}(z^j)}{2} - 2C \left(\sum_{j=1}^{m} \|e_j\|^2 \right)^{1/2} - O(1) \frac{1}{\sqrt{N}} \|\gamma^*\| \left(\sum_{s=1}^{p^*} \|V_s\|^2 \right)^{1/2} \\
= & C^2 \min_j \frac{\Lambda_{min}(z^j)}{2} - 2C \left(\sum_{j=1}^{m} \|e_j\|^2 \right)^{1/2} - O(1)C, \qquad \text{(E.5)}
\end{aligned}
$$

where we have used that $\frac{1}{\sqrt{N}}\|\gamma^*\| = O(1)$ by the condition in the body of this theorem and $\|e_j\| = O_P(1)$ uniformly in j. Notice that $\frac{1}{2}C^2 \min_j \Lambda_{min}(z^j)$ is a quadratic function in C while the rest terms on RHS of (E.5) are linear in C. Since C can be sufficiently large, it is easy to know that RHS of (E.5) is positive with probability arbitrarily close to 1. The proof for (E.3) is now complete. ■

2). For simplicity, we show that $\Pr(\|\hat{b}_{\gamma,p}\| = 0) \to 1$ only. The proofs for $\hat{b}_{\gamma,j}$ with $j = p^* + 1, \ldots, p-1$ are the same. If $\|\hat{b}_{\gamma,p}\| \neq 0$, $\hat{B}$ must satisfy the following equation

$$
0 = \frac{\partial}{\partial b_p} Q_\gamma(B) = A_1 + A_2, \qquad \text{(E.6)}
$$

where $A_1 = -\sum_{i=1}^N 2X_{i,p}\left((Y_i - X_i'\hat{\beta}_{\gamma,1})L(Z_i, z^1, \hat{\Theta}), \ldots, (Y_i - X_i'\hat{\beta}_{\gamma,m})L(Z_i, z^m, \hat{\Theta})\right)'$ and $A_2 = \frac{\gamma_p}{\|b_p\|} b_p$. For $s = 1, \ldots, m$, we can further write each element of A_1 as follows.

$$
\begin{aligned}
\frac{1}{\sqrt{N}} A_{1,s} &= -\frac{1}{\sqrt{N}} \sum_{i=1}^{N} 2X_{i,p} \left(X_i'(\beta_0(\bar{Z}_i) - \hat{\beta}_{\gamma,s}) + \varepsilon_i \right) L(Z_i, z^s, \hat{\Theta}) \\
&= -\frac{1}{\sqrt{N}} \sum_{i=1}^{N} 2X_{i,p}X_i'(\beta_0(\bar{Z}_i) - \hat{\beta}_{\gamma,s})L(Z_i, z^s, \hat{\Theta}) - \frac{1}{\sqrt{N}} \sum_{i=1}^{N} 2X_{i,p}\varepsilon_i L(Z_i, z^s, \hat{\Theta}) \\
&= -\frac{1}{\sqrt{N}} \sum_{i=1}^{N} 2X_{i,p}X_i'(\beta_0(\bar{Z}_i) - \beta_0(z^s))L(Z_i, z^s, \hat{\Theta}) \\
&\quad -\frac{1}{\sqrt{N}} \sum_{i=1}^{N} 2X_{i,p}X_i'(\beta_0(z^s) - \hat{\beta}_{\gamma,s})L(Z_i, z^s, \hat{\Theta}) + O_P(1)
\end{aligned}
$$

$$= O_P(1),$$

where the third equality follows from (3) of Lemma E.1; the last equality follows from (E.4) and the first result of this theorem.

On the other hand, $\left\|\frac{1}{\sqrt{N}}A_2\right\| = \frac{1}{\sqrt{N}}\gamma_p \geq \frac{1}{\sqrt{N}}\min_{s\in\{p^*+1,\ldots,p\}}\gamma_s \geq \omega_2$ by the condition in the body of this theorem, where ω_2 is sufficiently large. Therefore, $\Pr(\|A_1\| < \|A_2\|) \to 1$, which implies that, with probability tending to 1, (E.6) does not hold. The above analysis implies that $\hat{b}_{\gamma,p}$ must be located at the place where the objective function (2.7) is not differentiable with respect to b_p. Since equation (2.7) of the main file is only not differentiable with respect to b_p at the origin, we immediately obtain that $\Pr(\|\hat{b}_{\gamma,p}\| = 0) \to 1$. The same procedure of proof applies to $\hat{b}_{\gamma,j}$ with $j = p^*+1,\ldots,p-1$. The proof is then complete. ■

Proof of Theorem 2.2:

By Theorem 2.1, we know that $\hat{\beta}_{\gamma,js} = 0$ for $j = 1,\ldots,m$ and $s = p^*+1,\ldots,p$ with probability tending to one. After some simple algebra, we can obtain the first derivative of equation (2.7) of the main file with respect to β_j for $j = 1,\ldots,m$. Then it is easy to know that $\hat{\beta}_{\gamma,jU}$ must be the solution of the following equation

$$\frac{2}{N}\sum_{i=1}^{N} X_{iU}\left(Y_i - X_{iU}'\hat{\beta}_{\gamma,jU}\right)L(Z_i, z^j, \hat{\Theta}) - \frac{1}{N}D\hat{\beta}_{\gamma,jU} = 0,$$

where $\hat{\beta}_{\gamma,jU} = (\hat{\beta}_{\gamma,j1},\ldots,\hat{\beta}_{\gamma,jp^*})'$ and $D = \operatorname{diag}\left(\gamma_1\|\hat{b}_{\gamma,1}\|^{-1},\ldots,\gamma_{p^*}\|\hat{b}_{\gamma,p^*}\|^{-1}\right)$. It implies that $\hat{\beta}_{\gamma,jU}$ must have the form

$$\hat{\beta}_{\gamma,jU} = \left(\frac{1}{N}\sum_{i=1}^{N} X_{iU}X_{iU}'L(Z_i, z^j, \hat{\Theta}) + \frac{1}{2N}D\right)^{-1}\frac{1}{N}\sum_{i=1}^{N} X_{iU}Y_iL(Z_i, z^j, \hat{\Theta}).$$

Comparing with the oracle estimator,

$$\left\|\hat{\beta}_{\gamma,jU} - \hat{\beta}_{ora}(\bar{z}^j)\right\| \leq \left\|\Sigma_N(z^j)\right\|\left\|\frac{1}{N}\sum_{i=1}^{N} X_{iU}Y_iL(Z_i, z^j, \hat{\Theta})\right\|, \tag{E.7}$$

where $\Sigma_N(z^j)$ is denoted as

$$\Sigma_N(z^j) = \left(\frac{1}{N}\sum_{i=1}^{N} X_{iU}X_{iU}'L(Z_i, z^j, \hat{\Theta}) + \frac{1}{2N}D\right)^{-1} - \left(\frac{1}{N}\sum_{i=1}^{N} X_{iU}X_{iU}'L(Z_i, z^j, \hat{\Theta})\right)^{-1}.$$

Since $\Sigma_N(z^j)$ has finite dimensions, it is easy to know that the rate of $\left\|\Sigma_N(z^j)\right\|$ converging to 0 is the same as

$$\left\|\frac{1}{N}\sum_{i=1}^{N} X_{iU}X_{iU}'L(Z_i, z^j, \hat{\Theta}) + \frac{1}{2N}D - \frac{1}{N}\sum_{i=1}^{N} X_{iU}X_{iU}'L(Z_i, z^j, \hat{\Theta})\right\| = \left\|\frac{1}{2N}D\right\| = O_P\left(\frac{\|\gamma^*\|}{N}\right).$$

Moreover, by (2) and (4) of Lemma E.1, $\frac{1}{N}\sum_{i=1}^{N} X_{iU}Y_iL(Z_i, z^j, \hat{\Theta}) = O_P(1)$. Therefore, for $j = 1,\ldots,m$, $\left\|\hat{\beta}_{\gamma,jU} - \hat{\beta}_{ora}(\bar{z}^j)\right\| = O_P\left(\frac{\|\gamma^*\|}{N}\right)$. The proof is now complete. ■

Proof of Theorem 2.3:

(1) For an arbitrary model S, we say it is under-fitted if it misses at least one variable with a nonzero coefficient (i.e. $S \subset \mathcal{A}^c$ but $\mathcal{A}^c \neq S$); it is over fitted if S covers all relevant variables but also includes at least one redundant regressor (i.e. $\mathcal{A}^c \subset S$ but $\mathcal{A}^c \neq S$). Then, according to whether the model S_γ is under fitted, correctly fitted, or over fitted, we create three mutually exclusive sets $A^- = \{\tilde{\gamma} \in \mathbb{R} : S_{\tilde{\gamma}} \subset \mathcal{A}^c, S_{\tilde{\gamma}} \neq \mathcal{A}^c\}$, $A^0 = \{\tilde{\gamma} \in \mathbb{R} : S_{\tilde{\gamma}} = \mathcal{A}^c\}$ and $A^+ = \{\tilde{\gamma} \in \mathbb{R} : S_{\tilde{\gamma}} \supset \mathcal{A}^c, S_{\tilde{\gamma}} \neq \mathcal{A}^c\}$. Suppose that $\tilde{\beta}_j$ for $j = 1, \ldots, m$ are the unregularized estimators and there is a sequence $\{\hat{\gamma}_N\}$ that ensures (2.10) of the main file satisfies the conditions required by Theorem 2.1. For example, say those used in the section of Monte Carlo study.

Case 1: Under-fitted model, i.e. $S \subset \mathcal{A}^c$ but $\mathcal{A}^c \neq S$. Without losing generality, we assume that only one variable is missing, so we assume that the first $p^* - 1$ elements of $\hat{\beta}_{\tilde{\gamma},j}$ are obtained from the under-fitted model and the rest $p - p^* + 1$ elements of $\hat{\beta}_{\tilde{\gamma},j}$ are 0.

We then write

$$\begin{aligned}
RSS_{\tilde{\gamma}} &= \frac{1}{N}\sum_{j=1}^{m}\sum_{i=1}^{N}\left(Y_i - X_i'\hat{\beta}_{\tilde{\gamma},j}\right)^2 L(Z_i, z^j, \hat{\Theta}) \\
&= \frac{1}{N}\sum_{j=1}^{m}\sum_{i=1}^{N}\left(Y_i - X_i'\tilde{\beta}_j + X_i'\tilde{\beta}_j - X_i'\hat{\beta}_{\tilde{\gamma},j}\right)^2 L(Z_i, z^j, \hat{\Theta}) \\
&= \frac{1}{N}\sum_{j=1}^{m}\sum_{i=1}^{N}\left(Y_i - X_i'\tilde{\beta}_j\right)^2 L(Z_i, z^j, \hat{\Theta}) + \frac{1}{N}\sum_{j=1}^{m}\sum_{i=1}^{N}\left(X_i'\tilde{\beta}_j - X_i'\hat{\beta}_{\tilde{\gamma},j}\right)^2 L(Z_i, z^j, \hat{\Theta}) \\
&\quad + \frac{2}{N}\sum_{j=1}^{m}\sum_{i=1}^{N}\left(\tilde{\beta}_j - \hat{\beta}_{\tilde{\gamma},j}\right)' X_i\left(Y_i - X_i'\tilde{\beta}_j\right) L(Z_i, z^j, \hat{\Theta}) \\
&= \frac{1}{N}\sum_{j=1}^{m}\sum_{i=1}^{N}\left(Y_i - X_i'\tilde{\beta}_j\right)^2 L(Z_i, z^j, \hat{\Theta}) + \frac{1}{N}\sum_{j=1}^{m}\sum_{i=1}^{N}\left(X_i'\tilde{\beta}_j - X_i'\hat{\beta}_{\tilde{\gamma},j}\right)^2 L(Z_i, z^j, \hat{\Theta}) \\
&\equiv RSS^* + R_{2\tilde{\gamma}}
\end{aligned}$$

where the fourth equality is due to $\frac{2}{N}\sum_{j=1}^{m}\sum_{i=1}^{N}\left(\tilde{\beta}_j - \hat{\beta}_{\tilde{\gamma},j}\right)' X_i\left(Y_i - X_i'\tilde{\beta}_j\right) L(Z_i, z^j, \hat{\Theta}) = 0$ by the definition of unregularized estimators.

We now consider $R_{2\tilde{\gamma}}$ and write

$$\begin{aligned}
R_{2\tilde{\gamma}} &= \frac{1}{N}\sum_{j=1}^{m}\sum_{i=1}^{N}\left(\tilde{\beta}_j - \hat{\beta}_{\tilde{\gamma},j}\right)' X_i X_i' L(Z_i, z^j, \hat{\Theta})\left(\tilde{\beta}_j - \hat{\beta}_{\tilde{\gamma},j}\right) \\
&= \sum_{j=1}^{m}\left(\tilde{\beta}_j - \hat{\beta}_{\tilde{\gamma},j}\right)' \Sigma_1(z^j)\left(\tilde{\beta}_j - \hat{\beta}_{\tilde{\gamma},j}\right) + O_P\left(\frac{1}{\sqrt{N}}\right) \\
&\geq \sum_{j=1}^{m}\Lambda_{min}(z^j)\left\|\tilde{\beta}_j - \hat{\beta}_{\tilde{\gamma},j}\right\|^2 + O_P\left(\frac{1}{\sqrt{N}}\right) \\
&= O(1)\sum_{j=1}^{m}\left\|\tilde{\beta}_j - \hat{\beta}_{\tilde{\gamma},j}\right\|^2 + O_P\left(\frac{1}{\sqrt{N}}\right) \geq O(1)\sum_{j=1}^{m}\tilde{\beta}_{j,p^*}^2 + O_P\left(\frac{1}{\sqrt{N}}\right),
\end{aligned}$$

where $\Sigma_1(z^j) = \Pr(\bar{z}^j)E[X_iX_i'|\bar{z}^j]\Delta_1(\tilde{z}^j)$ and $\Delta_1(\tilde{z}^j)$ is denoted in (E.2); $\Lambda_{min}(z^j)$ denotes the minimum eigenvalue of $\Pr(\bar{z}^j)E[X_iX_i'|\bar{z}^j]\Delta_1(\tilde{z}^j)$; $\tilde{\beta}_{j,p^*}$ denotes the p^*th element of $\tilde{\beta}_j$; the second equality follows from (1) of Lemma E.1 of the Appendix; the first inequality follows from Assumption 3.

Similarly, we can obtain that $RSS_{\hat{\gamma}_N} \equiv RSS^* + R_{2\hat{\gamma}_N}$, where

$$
\begin{aligned}
R_{2\hat{\gamma}_N} &= \frac{1}{N}\sum_{j=1}^{m}\sum_{i=1}^{N}\left(\tilde{\beta}_j - \hat{\beta}_{\hat{\gamma}_N,j}\right)' X_i X_i' L(Z_i, z^j, \hat{\Theta})\left(\tilde{\beta}_j - \hat{\beta}_{\hat{\gamma}_N,j}\right) \\
&= \sum_{j=1}^{m}\left(\tilde{\beta}_j - \hat{\beta}_{\hat{\gamma}_N,j}\right)' \Sigma_1(z^j)\left(\tilde{\beta}_j - \hat{\beta}_{\hat{\gamma}_N,j}\right) + O_P\left(\frac{1}{\sqrt{N}}\right) \\
&\le \sum_{j=1}^{m}\Lambda_{max}(z^j)\left\|\tilde{\beta}_j - \hat{\beta}_{\hat{\gamma}_N,j}\right\|^2 + O_P\left(\frac{1}{\sqrt{N}}\right) \\
&\le O(1)\sum_{j=1}^{m}\left\|\tilde{\beta}_j - \hat{\beta}_{\hat{\gamma}_N,j}\right\|^2 + O_P\left(\frac{1}{\sqrt{N}}\right) \\
&\le O(1)\sum_{j=1}^{m}\left\|\tilde{\beta}_j - \beta_0(\bar{z}^j)\right\|^2 + O(1)\sum_{j=1}^{m}\left\|\hat{\beta}_{\hat{\gamma}_N,j} - \beta_0(\bar{z}^j)\right\|^2 + O_P\left(\frac{1}{\sqrt{N}}\right) = o_P(1),
\end{aligned}
$$

where $\Lambda_{max}(z^j)$ denotes the maximum eigenvalue of $\Pr(\bar{z}^j)E[X_iX_i'|\bar{z}^j]\Delta_1(\tilde{z}^j)$; the second equality follows from (1) of Lemma E.1 of the Appendix; the first inequality follows from Assumption 3; the last equality follows from using result (1) of Theorem 2.1 on $\sum_{j=1}^m \left\|\tilde{\beta}_j - \beta_0(\bar{z}^j)\right\|^2$ and using results (1)-(2) of Theorem 2.1 on $\sum_{j=1}^m \left\|\hat{\beta}_{\hat{\gamma}_N,j} - \beta_0(\bar{z}^j)\right\|^2$.

Note by (5) of Lemma E.1 we can obtain $RSS^* \to_P \sum_{j=1}^m \sigma_\varepsilon^2(\bar{z}^j)\Pr(\bar{z}^j)\Delta_1(\tilde{z}^j)$. Based on the analysis on $R_{2\tilde{\gamma}}$ and $R_{2\hat{\gamma}_N}$, we then can further conclude that

$$
\Pr\left(\inf_{\tilde{\gamma}\in A^-} BIC_{\tilde{\gamma}} > BIC_{\hat{\gamma}_N}\right) \to 1.
$$

Case 2: Over-fitted model, i.e. $S \supset \mathcal{A}^c$ but $\mathcal{A}^c \neq S$. Consider $\forall\tilde{\gamma} \in A^+$ and recall that $\hat{B}_{\tilde{\gamma}}$ determines a model $S_{\tilde{\gamma}}$. Under such a model $S_{\tilde{\gamma}}$, we can define another unregularized estimate $\check{B}_{\tilde{\gamma}}$ as

$$
\check{B}_{\tilde{\gamma}} = \underset{\beta_1,\ldots,\beta_m}{\operatorname{argmin}} \frac{1}{N}\sum_{j=1}^{m}\sum_{i=1}^{N}\left(Y_i - X_i'\beta_j\right)^2 L(Z_i, z^j, \hat{\Theta}),
$$

where, for $j = 1,\ldots,m$, $\|\beta_{j,s}\| = 0$ with $\forall s \notin S_{\tilde{\gamma}}$. In other words, $\check{B}_{\tilde{\gamma}} = (\check{\beta}_1,\ldots,\check{\beta}_m)'$ is the unregularized estimator under the model determined by $\hat{B}_{\tilde{\gamma}}$. By definition, we obtain immediately that $RRS_{\tilde{\gamma}} \ge RRS_{S_{\tilde{\gamma}}}$, where

$$
RRS_{S_{\tilde{\gamma}}} = \frac{1}{N}\sum_{j=1}^{m}\sum_{i=1}^{N}\left(Y_i - X_i'\check{\beta}_j\right)^2 L(Z_i, z^j, \hat{\Theta}).
$$

It follows that

$$
\begin{aligned}
&\ln RRS_{\tilde{\gamma}} - \ln RRS^* \ge \ln RRS_{S_{\tilde{\gamma}}} - \ln RRS^* \\
&= \ln\left\{\frac{RRS^*}{RRS^*} + \frac{1}{N\cdot RRS^*}\sum_{j=1}^{m}\sum_{i=1}^{N}\left(\tilde{\beta}_j - \check{\beta}_j\right)' X_iX_i'L(Z_i, z^j, \hat{\Theta})\left(\tilde{\beta}_j - \check{\beta}_j\right)\right\} \\
&\ge -\frac{O(1)}{N\cdot RRS^*}\sum_{j=1}^{m}\sum_{i=1}^{N}\left(\tilde{\beta}_j - \check{\beta}_j\right)' X_iX_i'L(Z_i, z^j, \hat{\Theta})\left(\tilde{\beta}_j - \check{\beta}_j\right)
\end{aligned}
$$

$$\geq -\frac{O_P(1)}{RRS^*}\sum_{j=1}^{m}\Lambda_{max}(z^j)\left\|\tilde{\beta}_j-\breve{\beta}_j\right\|^2$$

$$\geq -\frac{O_P(1)}{RRS^*}\sum_{j=1}^{m}\Lambda_{max}(z^j)\left\|\tilde{\beta}_j-\beta_0(\bar{z}^j)\right\|^2-\frac{O_P(1)}{RRS^*}\sum_{j=1}^{m}\Lambda_{max}(z^j)\left\|\beta_0(\bar{z}^j)-\breve{\beta}_j\right\|^2\geq -\left|O_P\left(\frac{1}{N}\right)\right|,$$

where $\Lambda_{max}(z^j)$ denotes the maximum eigenvalue of $\Pr(\bar{z}^j)E[X_iX_i'|\bar{z}^j]\Delta_1(\bar{z}^j)$; $\tilde{\beta}_j$ for $j=1,\ldots,m$ are the same unregularized estimators as those used in **Case 1**; the second inequality follows from (1) of Lemma E.1 and Assumption 3; the fourth inequality follows from using result (1) of Theorem 2.1 on both $\sum_{j=1}^{m}\Lambda_{max}(z^j)\left\|\tilde{\beta}_j-\beta_0(\bar{z}^j)\right\|^2$ and $\sum_{j=1}^{m}\Lambda_{max}(z^j)\left\|\beta_0(\bar{z}^j)-\breve{\beta}_j\right\|^2$.

Similarly, we can obtain that $\ln RRS_{\hat{\gamma}_N}-\ln RRS^*=O_P\left(\frac{1}{N}\right)$. Thus, we obtain that

$$\ln RRS_{\tilde{\gamma}}-\ln RRS_{\hat{\gamma}_N}\geq -\left|O_P\left(\frac{1}{N}\right)\right|.$$

We then write

$$\inf_{\tilde{\gamma}\in A^+}BIC_{\tilde{\gamma}}-BIC_{\hat{\gamma}_N}=\ln RRS_{\tilde{\gamma}}-\ln RRS_{\hat{\gamma}_N}+(df_{\tilde{\gamma}}-df_{\hat{\gamma}_N})\frac{\ln N}{N}.$$

By Theorem 2.1, we know that $\Pr(df_{\hat{\gamma}_N}\to p^*)=1$. Since $\tilde{\gamma}\in A^+$, we must have that $\Pr(df_{\tilde{\gamma}}\geq p^*+1)\to 1$. Then it is clear

$$\Pr\left(\inf_{\tilde{\gamma}\in A^+}BIC_{\tilde{\gamma}}>BIC_{\hat{\gamma}_N}\right)\to 1.$$

Combining Cases 1 and 2, we obtain that $\Pr\left(\inf_{\tilde{\gamma}\in A^-\cup A^+}BIC_{\tilde{\gamma}}>BIC_{\hat{\gamma}_N}\right)\to 1$. It further indicates that $\Pr\left(S_{\hat{\hat{\gamma}}}\to\mathcal{A}^c\right)=1$. It then completes the proof.

(2-3) The second and third results of this theorem follows by noticing that setting $\tilde{\gamma}$ to a large constant satisfies all the conditions required by Theorem 2.2 and the first result of this theorem. Thus, we have

$$\hat{\beta}_{\hat{\hat{\gamma}},jU}-\beta_0(\bar{z}^j)=\hat{\beta}_{ora}(\bar{z}^j)-\beta_0(\bar{z}^j)+O_P\left(\frac{1}{N}\right).$$

Then the results follow from Theorems 2 and 4 of Li et al. (2013). ■

References

Aitchison, J. and Aitken, C. (1976), 'Multivariate binary discrimination by the kernel method', *Biometrika* **63**(3), 413–420.

Antecol, H. and Bedard, K. (2006), 'Unhealthy assimilation: Why do immigrants converge to american health status levels ?', *Demography* **43**(2), 337–360.

Au, N., Hauck, K. and Hollingsworth, B. (2013), 'Employment, work hours and weight gain among middle–aged women', *International Journal of Obesity* **37**(5), 718–724.

Cawley, J. (2011), *The Oxford Handbook of the Social Science of Obesity*, Oxford University Press.

Cawley, J. and Scholder, S. v. H. K. (2013), The demand for cigarettes as derived from the demand for weight control, Technical report, National Bureau of Economic Research.

Chan, J. C. and Tobias, J. L. (2015), 'Priors and posterior computation in linear endogenous variable models with imperfect instruments', *Journal of Applied Econometrics* **30**(4), 650–674.

Chen, J., Gao, J. and Li, D. (2012), 'Semiparametric trending panel data models with cross–sectional dependence', *Journal of Econometrics* **171**(1), 71–85.

Cohen, A. K., Rai, M., Rehkopf, D. H. and Abrams, B. (2013), 'Educational attainment and obesity: a systematic review', *Obesity Reviews* **14**(12), 989–1005.

Colditz, G. A., Giovannucci, E., Rimm, E. B., Stampfer, M. J., Rosner, B., Speizer, F. E., Gordis, E. and Willett, W. C. (1991), 'Alcohol intake in relation to diet and obesity in women and men', *American Journal of Clinical Nutrition* **54**(1), 49–55.

Dardanoni, V., Modica, S. and Peracchi, F. (2011), 'Regression with imputed covariates: A generalized missing-indicator approach', *Journal of Econometrics* **162**(2), 362–368.

Delgado, M. S. (2013), 'A smooth coefficient model of carbon emissions', *Empirical Economics* **45**(3), 1049–1071.

El-Sayed, A. M., Scarborough, P. and Galea, S. (2012), 'Unevenly distributed: a systematic review of the health literature about socioeconomic inequalities in adult obesity in the united kingdom', *BMC Public Health* **12**(1), 18.

Faith, M., Butryn, M., Wadden, T., Fabricatore, A., Nguyen, A. and Heymsfield, S. (2011), 'Evidence for prospective associations among depression and obesity in population-based studies', *Obesity Reviews* **12**(5), e438–e453.

Fan, J. and Li, R. (2001), 'Variable selection via nonconcave penalized likelihood and its oracle properties', *Journal of the American Statistical Association* **96**(456), 1348–1360.

Fan, J. and Zhang, W. (1999), 'Statistical estimation in varying coefficient models', *Annals of Statistics* **27**(5), 1491–1518.

Fontaine, K. R., Redden, D. T., Wang, C., Westfall, A. O. and Allison, D. B. (2003), 'Years of life lost due to obesity', *JAMA* **289**(2), 187–193.

Galani, C. and Schneider, H. (2007), 'Prevention and treatment of obesity with lifestyle interventions: review and meta-analysis', *International Journal of Public Health* **52**(6), 348–359.

Gao, J. and Phillips, P. (2013), 'Functional coefficient nonstationary regression with non– and semi–parametric cointegration', *Available at https://ideas.repec.org/p/msh/ebswps/2013-16.html* .

Hastie, T. and Tibshirani, R. (1993), 'Varying-coefficient models', *Journal of the Royal Statistical Society. Series B (Methodological)* **55**(4), 757–796.

Hendricks, N. P. and Smith, A. (2015), 'Grouped coefficients to reduce bias in heterogeneous dynamic panel models with small t', *Applied Economics* **47**(40), 4335–4348.

Hill, J. O. and Peters, J. C. (1998), 'Environmental contributions to the obesity epidemic', *Science* **280**(5368), 1371–1374.

Hu, Y. and Sasaki, Y. (2015), 'Closed-form estimation of nonparametric models with non-classical measurement errors', *Journal of Econometrics* **185**(2), 392–408.

Kelly, I. R. and Markowitz, S. (2009), 'Incentives in obesity and health insurance', *Inquiry* **46**(4), 418–432.

Kline, B. and Tobias, J. L. (2008), 'The wages of bmi: Bayesian analysis of a skewed treatment–response model with nonparametric endogeneity', *Journal of Applied Econometrics* **23**(6), 767–793.

Li, Q., Ouyang, D. and Racine, J. S. (2013), 'Categorical semiparametrics varying–coefficient models', *Journal of Applied Econometrics* **28**(4), 551–579.

Li, Q. and Racine, J. S. (2010), 'Smooth varying–coefficient estimation and inference for qualitative and quantitative data', *Econometric Theory* **26**(6), 1607–1637.

Lipowicz, A., Gronkiewicz, S. and Malina, R. M. (2002), 'Body mass index, overweight and obesity in married and never married men and women in poland', *American Journal of Human Biology* **14**(4), 468–475.

Lounici, K., Pontil, M., Van De Geer, S. and Tsybakov, A. B. (2011), 'Oracle inequalities and optimal inference under group sparsity', *Annals of Statistics* **39**(4), 2164–2204.

Millimet, D. L. and Tchernis, R. (2013), 'Estimation of treatment effects without an exclusion restriction: with an application to the analysis of the school breakfast program', *Journal of Applied Econometrics* **28**(6), 982–1017.

Mukherji, A., Roychoudhury, S., Ghosh, P. and Brown, S. (n.d.), 'Etimating health demand for an ageing population: a felxible and robust bayesian joint model', *Forthcoming in Journal of Applied Econometrics. Available at DOI: 10.1002/jae.2463* .

Oza-Frank, R. and Cunningham, S. A. (2010), 'The weight of us residence among immigrants: a systematic review', *Obesity Reviews* **11**(4), 271–280.

Sobal, J., Rauschenbach, B. S. and Frongillo, E. A. (1992), 'Marital status, fatness and obesity', *Social Science & Medicine* **35**(7), 915–923.

Wang, H. and Leng, C. (2007), 'Unified lasso estimation by least squares approximation', *Journal of the American Statistical Association* **102**(479), 1039–1048.

Wang, H. and Xia, Y. (2009), 'Shrinkage estimation of the varying coefficient', *Journal of the American Statistical Association* **104**(486), 747—757.

Wang, L., Li, H. and Huang, J. Z. (2008), 'Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements', *Journal of the American Statistical Association* **103**(484), 1556–1569.

WHO (2015), 'Obesity and overweight fact sheet no 311', http://www.who.int/mediacentre/factsheets/fs311/en/.

Yu, Y. (2012), 'Educational differences in obesity in the united states: a closer look at the trends', *Obesity* **20**(4), 904–908.

Yuan, M. and Lin, Y. (2006), 'Model selection and estimation in regression with grouped variables', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**(1), 49–67.

Zhang, Q. and Wang, Y. (2004), 'Socioeconomic inequality of obesity in the united states: do gender, age, and ethnicity matter?', *Social Science & Medicine* **58**(6), 1171–1180.