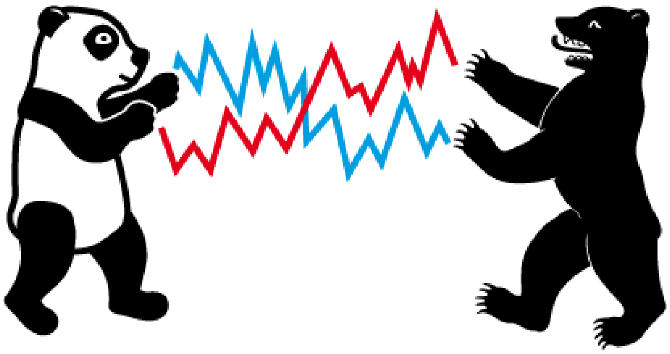
IRTG 1792 Discussion Paper 2018-052

Nonparametric Additive Instrumental Variable Estimator: A Group Shrinkage Estimation Perspective

Qingliang Fan *
Wei Zhong *



* Xiamen University, People's Republic of China

This research was supported by the Deutsche Forschungsgemeinschaft through the International Research Training Group 1792 "High Dimensional Nonstationary Time Series".

http://irtg1792.hu-berlin.de
ISSN 2568-5619

International Research Training Group 1792

Taylor & Francis
Taylor & Francis Group

Journal of Business & Economic Statistics

ISSN: 0735-0015 (Print) 1537-2707 (Online) Journal homepage: http://www.tandfonline.com/loi/ubes20

Nonparametric Additive Instrumental Variable Estimator: A Group Shrinkage Estimation Perspective

Qingliang Fan & Wei Zhong

To cite this article: Qingliang Fan & Wei Zhong (2018) Nonparametric Additive Instrumental Variable Estimator: A Group Shrinkage Estimation Perspective, Journal of Business & Economic Statistics, 36:3, 388-399, DOI: 10.1080/07350015.2016.1180991

To link to this article: https://doi.org/10.1080/07350015.2016.1180991

Accepted author version posted online: 27 Apr 2016.
Published online: 28 Apr 2017.

Nonparametric Additive Instrumental Variable Estimator: A Group Shrinkage Estimation Perspective

Qingliang FAN and Wei ZHONG

Wang Yanan Institute for Studies in Economics (WISE), Department of Statistics, School of Economics and Fujian Key Laboratory of Statistical Science, Xiamen University, Fujian, China (*michaelqfan@gmail.com; wzhong@xmu.edu.cn*)

In this article, we study a nonparametric approach regarding a general nonlinear reduced form equation to achieve a better approximation of the optimal instrument. Accordingly, we propose the nonparametric additive instrumental variable estimator (NAIVE) with the adaptive group Lasso. We theoretically demonstrate that the proposed estimator is root-n consistent and asymptotically normal. The adaptive group Lasso helps us select the valid instruments while the dimensionality of potential instrumental variables is allowed to be greater than the sample size. In practice, the degree and knots of B-spline series are selected by minimizing the BIC or EBIC criteria for each nonparametric additive component in the reduced form equation. In Monte Carlo simulations, we show that the NAIVE has the same performance as the linear instrumental variable (IV) estimator for the truly linear reduced form equation. On the other hand, the NAIVE performs much better in terms of bias and mean squared errors compared to other alternative estimators under the high-dimensional nonlinear reduced form equation. We further illustrate our method in an empirical study of international trade and growth. Our findings provide a stronger evidence that international trade has a significant positive effect on economic growth.

KEY WORDS: Adaptive group Lasso; Instrumental variables; Nonparametric additive model; Optimal estimator; Variable selection.

1. INTRODUCTION

The instrumental variable (IV) method is a signature technique in econometrics. The method has broad applications in various empirical studies when there are endogeneity issues in the structural equation. For example, Angrist and Krueger (1991) used quarter of birth as the instrumental variable for years of education in the study of returns to education. The validity of the inference using the IV method was studied by, for example, Hahn and Hausman (2002) and Berkowitz et al. (2012). Bound, Jaeger, and Baker (1995) showed that when instruments are weak, a mild violation of exogeneity could result in large inconsistency of the IV estimator. In empirical works, if we enforced the exogeneity condition in the searching for instruments, we would often have instruments that are only weakly correlated with the endogenous variable (Staiger and Stock 1997). IV inference with many possibly weak instruments is common, but the inference may have poor properties (Chao and Swanson 2005; Andrews, Moreira, and Stock 2006; Hausman et al. 2012). Hansen, Hausman, and Newey (2008) proposed corrected standard errors for estimation using many instruments. In practice, the first-stage F-statistic "rule of thumb" is suitable only for low-dimensional linear models (when the number of instruments is fixed and much smaller than the sample size) and it cannot identify which instrument is weak. In theory, the inclusion of many exogenous instruments can improve the efficiency of IV estimator. However, in practice, the IV estimator has higher bias if there are many weak instruments in the model. Hence, the choice of the set of instruments is crucial for the finite sample property of IV estimator (Donald and Newey 2001). The optimality of IV method requires the estimation of the conditional expectation (Amemiya 1974). It thus motivated us to develop a method that would allow us to select from a very large set of candidate instruments that can better estimate the conditional expectation.

The structural equation, while frequently being studied in either parametric or nonparametric framework (Newey and Powell 2003; Ozabaci, Henderson, and Su 2014), is often based on economic theory, for example, trade theory (Melitz 2003). The reduced form model, on the other hand, is more data-driven, and it is usually not guided by certain economic theory that suggests covariates to be included. For a given set of instruments, the optimal instrument often involves conditional expectations of nonlinear functions of endogenous variables. It is thus important to assume general nonlinear reduced form equations (Newey 1990). A linear reduced form equation is often assumed in empirical studies due to its simplicity even when the true relationship is nonlinear. For instance, Behrman et al. (2012) studied the causal effect of financial literacy on wealth accumulation using many instruments, including individual education environment variables, macroeconomic condition variables, and family background variables, among others. For a better approximation of the optimal instruments, the nonlinear reduced form might be more useful here. A good econometric procedure should be able to explore the unspecified nonlinear form while achieving the efficiency bound of IV estimator as well as better finite sam-

Journal of Business & Economic Statistics
July 2018, Vol. 36, No. 3
DOI: 10.1080/07350015.2016.1180991
Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/jbes*.

ple properties. Our simulation studies showed that the two-stage least squares (2SLS) estimator is more biased and inefficient if the nonlinear relationship exists in the reduced form model. This observation motivated us to study a nonparametric reduced form IV estimator, which should be widely applicable in practice.

In this article, we study high-dimensional nonparametric additive reduced form models with a large number of original observable instrumental variables, which can be larger than the sample size. Many researchers have studied additive nonparametric models (Linton et al. 1997). The nonparametric study of reduced form equation is often troubled by the curse of dimensionality (Newey 1990), which has been the focus of a substantial body of recent literature on high-dimensional problems. These high-dimensional methods include the Lasso (Tibshirani 1996), SCAD (Fan and Li 2001), group Lasso (Yuan and Lin 2006), adaptive Lasso (Zou 2006), adaptive group Lasso (Huang, Horowitz, and Wei 2010), and Dantzig selector (Candes and Tao 2007), among others. Recently, some studies have addressed high-dimensional endogeneity problem. Belloni et al. (2012) extended IV estimation to high dimension with heteroscedastic and non-Gaussian random disturbances using a modified Lasso method. Related ideas also appeared in Bai and Ng (2010), Carrasco (2012), Fan and Liao (2014), and Caner and Fan (2015). Lin et al. (2015) studied high-dimensional endogenous issue with applications in genomics. Our article differs from the aforementioned studies in two aspects. First, we focus on the high dimensionality of original economic variable observed by the researcher rather than the functions of those original observable variables. Second, we use the adaptive group Lasso approach to select additive components of instruments, such that the nonlinear relationship in reduced form equation is captured.

We assume a nonparametric additive reduced form equation to approximate the optimal instruments and propose the nonparametric additive instrumental variable estimator (NAIVE) with the adaptive group Lasso. We theoretically demonstrate that the proposed estimator is root-n consistent and asymptotically normal with optimal variance of IV estimator (Amemiya 1974; Chamberlain 1987; Newey 1990). The adaptive group Lasso shrinkage method selects the strong instruments consistently while the dimensionality of potential instruments is allowed to be greater than the sample size. In simulations, we show that the proposed estimator has smaller biases and mean squared errors compared to alternative methods among various model settings. It is worth noting that we use the BIC or EBIC criteria to choose the degree and knots of B-spline series for each nonparametric component. For a truly linear reduced form model, the proposed NAIVE automatically adopts the linearity form. Thus, our proposed method nests the linear reduced form as a special case, and its performance is the same as that of 2SLS in this case. Our method can be widely applicable, especially when the empirical researchers do not know which specific instruments should be included or the functional form of reduced models.

The article is organized as follows. In Section 2, we describe the methodology and present the nonparametric additive instrumental variable estimator (NAIVE) with the adaptive group Lasso. Section 3 presents the theoretical results. Section 4 shows the finite sample performance of our proposed estimator using Monte Carlo experiments. In Section 5, we illustrate our method in an empirical study of international trade and growth. Section 6 concludes the article. All the proofs are relegated to the Appendix.

2. METHODOLOGY

2.1 Some Preliminaries

We consider the following structural equation,

$$y_i = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0 + \varepsilon_i, \tag{2.1}$$

where y_i is the ith response variable, $\mathbf{x}_i$ is $d \times 1$ vector of explanatory variables, and $\boldsymbol{\beta}_0$ is $d \times 1$ vector of true parameters, for $i = 1, 2, \ldots, n$, where n is the sample size. The model is a general linear model if $E(\varepsilon_i|\mathbf{x}_i) = 0$. However, the endogeneity problem in empirical economic studies is common, such that $E(\varepsilon_i|x_{i\ell}) \neq 0$ for some $1 \leq \ell \leq d$. Without loss of generality, we assume the first d_e variables $\{x_{i\ell}, \ell = 1, 2, \ldots, d_e, 1 \leq d_e \leq d\}$ are endogenous, where d_e is fixed.

To solve the endogeneity problem, instrumental variables are employed to obtain a consistent estimator of the population regression coefficient $\boldsymbol{\beta}_0$. In practice, the choice of instruments affects the properties of IV estimators. We model the reduced form equation by using many potential instruments without knowing which one is useful. The $p \times 1$ vector of instrumental variables is denoted by $\mathbf{z}_i = (z_{i1}, \ldots, z_{ip})$. We consider the standard assumptions here, $E(\varepsilon_i|\mathbf{z}_i) = 0$ for all $i = 1, \ldots, n$, and $E(\xi_{i\ell}|\mathbf{z}_i) = 0$, for all $i = 1, \ldots, n$ and $\ell = 1, \ldots, d_e$, which is the exclusion restriction on the instrumental variables. We are interested in estimating the optimal conditional expectation $\mathbf{D}_i = D(\mathbf{z}_i) = E(\mathbf{x}_i|\mathbf{z}_i)$, which minimizes the asymptotic variance of IV estimator (Amemiya 1974). In many empirical economic studies, the linear reduced form model is often assumed to predict the endogenous variables $\mathbf{x}_i$ using $\mathbf{z}_i$. In the IV literature, the so-called k-class estimator is

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}(\mathbf{I} - k\mathbf{M})\mathbf{X})^{-1}(\mathbf{X}^{\mathrm{T}}(\mathbf{I} - k\mathbf{M})\mathbf{Y}), \tag{2.2}$$

where $\mathbf{M} = \mathbf{I} - \mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}$, and $\mathbf{X}$ is a $n \times d$ matrix formed by stacking the $\mathbf{x}_i$, $\mathbf{Z}$ is a $n \times p$ matrix formed by stacking the $\mathbf{z}_i$, $\mathbf{Y}$ is a $n \times 1$ matrix formed by stacking the y_i, and k is specified by the researcher. When $k = 0$, we have the OLS. When $k = 1$, it is the 2SLS. The limited information maximum likelihood (LIML) estimator and the Fuller's estimators are also well-known special cases of this k-class estimator.

In practice, many potential instruments, including their series terms, may be recruited to approximate the optimal instrument and improve the precision of IV estimators. On the other hand, if many irrelevant instruments are contained in the reduced form equation, the approximation of the optimal instrument is generally unsatisfactory and the IV estimator is less efficient. In some cases where the dimensionality of $\mathbf{z}_i$ is even higher than the sample size, the linear IV method fails. To address these issues, the model sparsity is usually assumed and the penalized approaches can be applied to improve the efficiency of IV estimators. Following the line of Belloni et al. (2012), in this article, we propose the first-stage parsimonious predictive models and estimated optimal instruments in IV models with potentially more instruments than the sample size n.

2.2 Nonparametric Additive Instrumental Variable Estimator

The performance of the linear IV estimator in the finite sample is largely dependent on the validity of linearity assumption. This phenomenon motivated us to consider a more general nonlinear reduced form equation to capture as much information of $\mathbf{x}_i$ as possible using instruments $\mathbf{z}_i$ under the high-dimensional model settings. This nonparametric idea for the reduced form model is consistent with Newey (1990). In this article, we consider the following nonparametric additive reduced form model with a large number of possible instruments. That is, for each $\ell = 1, \dots, d_e$,

$$x_{i\ell} = \mu_\ell + \sum_{j=1}^{p} f_{j\ell}(z_{ij}) + \xi_{i\ell}, \tag{2.3}$$

where μ_ℓ is the constant term, $f_{j\ell}(\cdot)$ is the jth unknown smooth univariate functions, and $\xi_{i\ell}$'s are iid random errors with mean 0 and finite variance. Here, the dimensionality p is potentially larger than the sample size n. For the model identification, we assume that all functions $f_{j\ell}(\cdot)$'s are centered, that is, $E[f_{j\ell}(z_j)] = 0$, $1 \le j \le p$, where z_j denotes the jth instrument. As it is more flexible and generally applicable than the ordinary linear model, the nonparametric additive reduced form model (2.3) could achieve a better approximation to the optimal instruments $E(\mathbf{x}_i|\mathbf{z}_i)$. Consequently, the IV estimator based on (2.3) is expected to be more efficient compared to the linear IV estimator. This conjecture will be confirmed both theoretically and numerically in the later sections.

To estimate the nonparametric components in (2.3), we use B-spline basis functions by following the idea of Huang, Horowitz, and Wei (2010). Let $\mathcal{S}_n$ be the space of polynomial splines of degrees $L \ge 1$ and let $\{\phi_k, k = 1, \dots, m_n\}$ be normalized B-spline basis functions for $\mathcal{S}_n$, where m_n is the sum of the polynomial degree L and the number of knots. Let $\psi_k(z_{ij}) = \phi_k(z_{ij}) - n^{-1}\sum_{i=1}^{n}\phi_k(z_{ij})$ be the centered B-spline basis functions for the jth instrument. Thus, for each $\ell = 1, \dots, d_e$, each $f_{nj\ell} \in \mathcal{S}_n$ can be represented by the linear combination of normalized B-spline series

$$f_{nj\ell}(z_{ij}) = \sum_{k=1}^{m_n} \gamma_{jk}\psi_k(z_{ij}), \quad 1 \le j \le p.$$

Under suitable smoothness conditions, the function $f_{j\ell}(z_{ij})$ in (2.3) can be well approximated by the function $f_{nj}(z_{ij})$ in $\mathcal{S}_n$ by carefully choosing the coefficients $\{\gamma_{j1}, \dots, \gamma_{jm_n}\}$ (Stone 1985). The model (2.3) can then be rewritten using an approximate linear reduced form

$$x_{i\ell} \approx \mu_\ell + \sum_{j=1}^{p}\sum_{k=1}^{m_n} \gamma_{jk\ell}\psi_k(z_{ij}) + \xi_{i\ell}. \tag{2.4}$$

Instead of directly estimating each nonlinear function $f_{j\ell}$ in (2.3), we are now allowed to estimate the parameter vector $\{\gamma_{j1\ell}, \dots, \gamma_{jm_n\ell}\}$ for each $\ell = 1, \dots, d_e$ using the OLS method when p is fixed and small.

To obtain a more precise IV estimator, we generally consider many potential instruments in practice to estimate the model (2.4). Under the high-dimensional reduced form models, if $p > n/m_n$, the OLS method fails to work for the model (2.4) due to the singularity of the design matrix. In high-dimensionality problems, the assumption that there exists only a small subset of valid instruments among many potential ones in the model (2.3) will be satisfied quite generally. We denote by $\mathcal{A}_\ell$ a set of instruments that are able to approximate the conditional expectation of the ℓth endogenous variable. That is, $f_{j\ell}(z) \ne 0$ for some z, $j \in \mathcal{A}_\ell$, but $f_{j\ell}(z) = 0$ for any z, $j \notin \mathcal{A}_\ell$, $\ell = 1, \dots, d_e$.

Let $\boldsymbol{\gamma}_{j\ell} = (\gamma_{j1\ell}, \dots, \gamma_{jm_n\ell})^{\mathrm{T}}$ be the $m_n \times 1$ vector of parameters corresponding to the jth instrument in (2.4) and denote $\boldsymbol{\gamma}_\ell = (\boldsymbol{\gamma}_{1\ell}^{\mathrm{T}}, \dots, \boldsymbol{\gamma}_{p\ell}^{\mathrm{T}})^{\mathrm{T}}$ by the $m_n p \times 1$ vector of parameters. Let $\mathbf{U}_{ij} = (\psi_1(z_{ij}), \dots, \psi_{m_n}(z_{ij}))^{\mathrm{T}}$ and $\mathbf{U}_j = (\mathbf{U}_{1j}, \dots, \mathbf{U}_{nj})^{\mathrm{T}}$ be the $n \times m_n$ design matrix for the jth instrument, and $\mathbf{U} = (\mathbf{U}_1, \dots, \mathbf{U}_p)$ be the corresponding $n \times m_n p$ design matrix. Let $\mathbf{X}_\ell = (x_{1\ell} - \bar{x}_\ell, \dots, x_{n\ell} - \bar{x}_\ell)^{\mathrm{T}}$ be the $n \times 1$ vector of ℓth centered endogenous variable for each $\ell = 1, \dots, d_e$. To select the significant instruments and estimate the component functions simultaneously, we consider the following penalized objective function with an adaptive group Lasso penalty (Huang, Horowitz, and Wei 2010) for each ℓth endogenous variable,

$$L_n(\boldsymbol{\gamma}_\ell; \lambda_n) = \|\mathbf{X}_\ell - \mathbf{U}\boldsymbol{\gamma}_\ell\|_2^2 + \lambda_n \sum_{j=1}^{p} \omega_{nj\ell}\|\boldsymbol{\gamma}_{j\ell}\|_2, \tag{2.5}$$

where $\|\cdot\|_2$ denotes the ℓ_2 norm of a vector, λ_n is a tuning parameter to control the shrinkage of parameters estimation, and $\omega_{nj\ell}$ is the positive weight for the jth group. We use the group Lasso estimator to obtain the weights (Huang, Horowitz, and Wei 2010) by defining

$$\omega_{nj\ell} = \begin{cases} \|\widetilde{\boldsymbol{\gamma}}_{j\ell}\|_2^{-1}, & \text{if } \|\widetilde{\boldsymbol{\gamma}}_{j\ell}\|_2 > 0, \\ \infty, & \text{if } \|\widetilde{\boldsymbol{\gamma}}_{j\ell}\|_2 = 0, \end{cases} \tag{2.6}$$

where $\widetilde{\boldsymbol{\gamma}}_{n\ell} = (\widetilde{\boldsymbol{\gamma}}_{1\ell}^{\mathrm{T}}, \widetilde{\boldsymbol{\gamma}}_{2\ell}^{\mathrm{T}}, \dots, \widetilde{\boldsymbol{\gamma}}_{p\ell}^{\mathrm{T}})^{\mathrm{T}}$ is the group Lasso estimator obtained by minimizing the penalized objective function with a group Lasso penalty, that is,

$$\begin{aligned} \widetilde{\boldsymbol{\gamma}}_{n\ell} &= \arg\min_{\boldsymbol{\gamma}_\ell} L_{n0}(\boldsymbol{\gamma}_\ell; \lambda_n) \\ &= \arg\min_{\boldsymbol{\gamma}_\ell} \left\{ \|\mathbf{X}_\ell - \mathbf{U}\boldsymbol{\gamma}_\ell\|_2^2 + \lambda_{n0} \sum_{j=1}^{p} \|\boldsymbol{\gamma}_{j\ell}\|_2 \right\}. \end{aligned} \tag{2.7}$$

The adaptive group Lasso estimator, via minimizing the objective function (2.5) with the weights in (2.6), is

$$\widehat{\boldsymbol{\gamma}}_{n\ell} = \left(\widehat{\boldsymbol{\gamma}}_{1\ell}^{\mathrm{T}}, \widehat{\boldsymbol{\gamma}}_{2\ell}^{\mathrm{T}}, \dots, \widehat{\boldsymbol{\gamma}}_{p\ell}^{\mathrm{T}}\right)^{\mathrm{T}} = \arg\min_{\boldsymbol{\gamma}_\ell} L_n(\boldsymbol{\gamma}_\ell; \lambda_n). \tag{2.8}$$

Denoting the selected set of instruments by $\widehat{\mathcal{A}}_\ell = \{j : \|\widehat{\boldsymbol{\gamma}}_{j\ell}\|_2 > 0\}$, the adaptive group Lasso estimators of μ_ℓ and $f_{j\ell}$ in (2.3) are

$$\widehat{\mu}_\ell = \frac{1}{n}\sum_{i=1}^{n} x_{i\ell}, \quad \widehat{f}_{nj\ell}(z_{ij}) = \sum_{k=1}^{m_n} \widehat{\gamma}_{jk\ell}\psi_k(z_{ij}), \quad j \in \widehat{\mathcal{A}}_\ell.$$

Then each ℓth endogenous variable, for $\ell = 1, \dots, d_e$, can be estimated accordingly by

$$\widehat{x}_{i\ell} = \widehat{\mu}_\ell + \sum_{j \in \widehat{\mathcal{A}}_\ell}\sum_{k=1}^{m_n} \widehat{\gamma}_{jk\ell}\psi_k(z_{ij}). \tag{2.9}$$

Denote $\widehat{\mathbf{x}}_i = (\widehat{x}_{i1}, \ldots, \widehat{x}_{id_e}, x_{id_e+1}, \ldots, x_{id})^{\mathrm{T}}$. Then the resulting IV estimator for $\boldsymbol{\beta}_0$ in the model (2.1) is

$$\widehat{\boldsymbol{\beta}} = \left(\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{x}}_i\mathbf{x}_i^{\mathrm{T}}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{x}}_i y_i\right). \tag{2.10}$$

We call $\widehat{\boldsymbol{\beta}}$ the nonparametric additive instrumental variable estimator (NAIVE) with the adaptive group Lasso. Because of the better approximation to the optimal instrument, it is shown both theoretically and numerically that the proposed NAIVE of $\boldsymbol{\beta}_0$ is root-n consistent and asymptotically normal.

In addition, we remark that the tuning parameters λ_{n0}, λ_n, and m_n are chosen adaptively in practice by minimizing the Bayesian information criterion (BIC) (Schwartz 1978; Wang, Li, and Tsai 2007) or extended BIC (EBIC) (Chen and Chen 2008) for (2.7) and (2.8). The BIC and EBIC are defined, respectively, by

$$\mathrm{BIC}(\lambda_n, m_n) = \log(\mathrm{RSS}_{\lambda_n}) + \hat{q}_\ell m_n \frac{\log n}{n},$$

$$\mathrm{EBIC}(\lambda_n, m_n) = \log(\mathrm{RSS}_{\lambda_n}) + \hat{q}_\ell m_n \frac{\log n + \nu \log p}{n},$$

where RSS_{λ_n} is the residual sum of squares for a given λ_n, $\hat{q}_\ell = |\widehat{\mathcal{A}}_\ell|$, the number of nonzero component functions for each $\ell = 1, \ldots, d_e$ and $0 \le \nu \le 1$ is a constant. When the nonparametric additive reduced form model (2.3) is indeed a linear model, that is, all component functions are linear, this data-driven approach can usually select $m_n = 1$. Subsequently, (2.3) degenerates to an ordinary linear model. In this case, the proposed NAIVE is the same as the linear 2SLS estimator if the true reduced form model is linear. Simulations in Section 4 confirm this result. In this sense, the NAIVE could be substantially useful for empirical studies to solve the endogeneity problem, especially when the relationship between the endogenous variables and instruments are unknown.

To briefly summarize, we present the following algorithm to find the NAIVE.

Algorithm Nonparametric Additive Instrumental Variable Estimator

Step 1. Obtain the group Lasso estimator $\widetilde{\boldsymbol{\gamma}}_{n\ell} = \arg\min_{\boldsymbol{\gamma}_\ell} L_{n0}(\boldsymbol{\gamma}_\ell; \lambda_n)$ in (2.7) where BIC or EBIC are applied to choose the tuning parameters λ_n and m_n.
Step 2. Define the weights based on (2.6) using the group Lasso estimator in Step 1and obtain the adaptive group Lasso estimator $\widehat{\boldsymbol{\gamma}}_{n\ell} = \arg\min_{\boldsymbol{\gamma}_\ell} L_n(\boldsymbol{\gamma}_\ell; \lambda_n)$ in (2.8).
Step 3. Estimate the fitted value of each ℓth endogenous variable in (2.9) and denote the new design matrix $\widehat{\mathrm{X}}_i = (\widehat{x}_{i1}, \ldots, \widehat{x}_{id_e}, x_{id_e+1}, \ldots, x_{id})^{\mathrm{T}}$.
Step 4. Regress the response y against $\widehat{\mathrm{X}}_i$ to obtain the NAIVE $\widehat{\boldsymbol{\beta}}$ based on (2.10).

3. THEORETICAL RESULTS

In this section, we present the theoretical results that the NAIVE with the adaptive group Lasso is root-n consistent and asymptotically normal. First, we assume some regularity conditions.

(C1) The support of each instrument z_j is $[a, b]$, where a and b are finite real numbers. The density function g_j of z_j in (2.3) satisfies $0 < K_1 \le g_j(z) \le K_2 < \infty$ on $[a, b]$ for $j = 1, \ldots, p$.

(C2) Let $\mathcal{F}$ be the class of functions f such that the rth derivative $f^{(r)}$ exists and satisfies a Lipschitz condition of order $\alpha \in (0, 1]$. That is,

$$\mathcal{F} = \left\{ f(\cdot) : |f^{(r)}(t_1) - f^{(r)}(t_2)| \le C|t_1 - t_2|^\alpha, \right.$$
$$\left. \text{for } t_1, t_2 \in [a, b] \text{ and a constant } C > 0 \right\},$$

where r is a nonnegative integer and $\alpha \in (0, 1]$ such that $s = r + \alpha > 1.5$. Suppose $f_{j\ell} \in \mathcal{F}$, $j = 1, \ldots, p$, $\ell = 1, \ldots, d_e$, in (2.3).

(C3) $\xi_{i\ell}$ satisfies the subexponential tail probability, $E[\exp(c|\xi_{i\ell}|)] < \infty$ for a finite positive constant c and $E(x_{i\ell}^2) < \infty$ for $i = 1, \ldots, n$, $\ell = 1, \ldots, d_e$. ε_i satisfies that $E(\varepsilon_i^3)$ is bounded away from zero and the infinity, $i = 1, \ldots, n$.

(C4) The number of the significant instruments $q_\ell = |\mathcal{A}_\ell|$ is fixed. There exists $c_\ell > 0$, such that $\min_{j \in \mathcal{A}_\ell} \|f_{j\ell}\|_2 \ge c_\ell$ for each $\ell = 1, \ldots, d_e$, where $\|f_{j\ell}\|_2^2 = \int_a^b f_{j\ell}^2(x)dx$.

All regularity conditions (C1)–(C4) are standard conditions for nonparametric estimation (Huang, Horowitz, and Wei 2010; Fan et al. 2011). In particular, in the nonparametric literature, Lipschitz Condition (C2) is commonly assumed to require that the function is smooth enough. Condition (C3) requires that the distribution of the random errors $\xi_{i\ell}$'s should not be too heavy-tailed, and it is satisfied for $\xi_{i\ell}$'s that are bounded uniformly or normally distributed.

Lemma 3.1. Using the group Lasso estimator $\widetilde{\boldsymbol{\gamma}}_{n\ell}$ with $\lambda_{n0} \asymp O(\sqrt{n}\log(m_n p))$ and $m_n \asymp O(n^{1/(2s+1)})$ to construct the weight for the adaptive group Lasso estimator. Suppose Conditions (C1)–(C4) hold, d_e is fixed, and $\lambda_n \asymp O(\sqrt{n})$, then

$$P(\widehat{\mathcal{A}}_\ell = \mathcal{A}_\ell) \to 1, \text{ as } n \to \infty, \text{ for } \ell = 1, \ldots, d_e. \tag{3.1}$$

$$\max_{1 \le \ell \le d_e} \sum_{j \in \mathcal{A}_\ell} \|\widehat{\boldsymbol{\gamma}}_{nj\ell} - \boldsymbol{\gamma}_{j\ell}\|_2^2 = O_p\left(n^{-(2s-1)/(2s+1)}\right). \tag{3.2}$$

$$\max_{1 \le \ell \le d_e} \sum_{j \in \mathcal{A}_\ell} \|\widehat{f}_{nj\ell} - f_{j\ell}\|_2^2 = O_p\left(n^{-2s/(2s+1)}\right). \tag{3.3}$$

This lemma shows the selection consistency of the adaptive group Lasso for high-dimensional nonparametric additive reduced form model. That is, the true set of significant instruments can be identified for each endogenous variable with probability tending to 1. It also establishes the estimation consistency of $\widehat{\boldsymbol{\gamma}}_{nj\ell}$ in the nonparametric additive reduced form model. Because the number of endogenous variables d_e is fixed, Lemma 3.1 essentially follows the results of theorem 3 in Huang, Horowitz, and Wei (2010). In addition, the consistency of the group Lasso estimator is the necessary condition for the results in Lemma 3.1. With $\lambda_{n0} \asymp O(\sqrt{n}\log(m_n p))$, theorem 1 of Huang, Horowitz, and Wei (2010) has shown that under conditions (C1)–(C4),

$$\|\widetilde{\boldsymbol{\gamma}}_{n\ell} - \boldsymbol{\gamma}_{\ell}\|_2^2 = \sum_{j=1}^{p} \|\widetilde{\boldsymbol{\gamma}}_{nj\ell} - \boldsymbol{\gamma}_{j\ell}\|_2^2 = O_p\left(\frac{m_n^2 \log(m_n p)}{n}\right) + O_p\left(\frac{m_n}{n}\right) + O\left(\frac{1}{m_n^{2s-1}}\right). \quad (3.4)$$

To ensure the consistency of $\widetilde{\boldsymbol{\gamma}}_{n\ell}$, the dimensionality of instruments is allowed to be $p = \exp(o(n/m_n^2)) = O(\exp(n^{\kappa}))$ with $0 < \kappa < 1 - 2/(2s+1)$, which can be much larger than the sample size n.

Denote by $D_{i\ell} = E(x_{i\ell}|\mathbf{z}_i) = \mu_\ell + \sum_{j\in\mathcal{A}_\ell} f_{j\ell}(z_{ij})$ the conditional expectation of ℓth endogenous variable given the instrumental variables, for $\ell = 1, \ldots, d_e$, so that the reduced form equation becomes $x_{i\ell} = D_{i\ell} + \xi_{i\ell}$. Let $\mathbf{D}_i = (D_{i1}, \ldots, D_{id_e}, x_{id_e+1}, \ldots, x_{id})^{\mathrm{T}}$. The following theorem presents the main result on the proposed NAIVE.

Theorem 3.1. Suppose that the structural disturbance is conditionally homoscedastic, that is, $\mathrm{Var}(\varepsilon_{il}|\mathbf{z}_i) = \sigma^2$. Under the regularity Conditions (C1)–(C4), the NAIVE with the adaptive group Lasso in (2.10) is $\sqrt{n}$−consistent and asymptotically normal. That is,

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \xrightarrow{d} N\left(0, \sigma^2\left\{E\left(\mathbf{D}_i\mathbf{D}_i^{\mathrm{T}}\right)\right\}^{-1}\right). \quad (3.5)$$

Theorem 3.1 demonstrates that the proposed NAIVE with the adaptive group Lasso is asymptotically unbiased, $\sqrt{n}$−consistent, and asymptotically normal. It also shows that our estimator achieves the optimal asymptotic variance $\sigma^2\{E(\mathbf{D}_i\mathbf{D}_i^{\mathrm{T}})\}^{-1}$ as discussed by Amemiya (1974), Chamberlain (1987), and Newey (1990). This result will be further manifested in the next simulation sections. Note that for the inference purpose, the asymptotic variance can be estimated by $\widehat{\sigma}^2(n^{-1}\sum_{i=1}^n \widehat{\mathbf{x}}_i\widehat{\mathbf{x}}_i^{\mathrm{T}})^{-1}$, where $\widehat{\sigma}^2 = n^{-1}\sum_{i=1}^n (y_i - \widehat{\mathbf{x}}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}})^2$. The theoretical result here is parallel with theorem 4 in Belloni et al. (2012) who studied the high-dimensional IV estimator with Lasso.

4. SIMULATIONS

In this section, we assess the finite sample performance of the proposed NAIVE method by Monte Carlo simulation studies. We consider a simple structural equation with one endogenous variable,

$$y_i = x_i\beta_0 + \varepsilon_i,$$

where y_i is the dependent variable of interest, x_i is the endogenous variable, $i = 1, 2, \ldots, n$, and the true coefficient $\beta_0 = 0.75$. The endogenous variable is generated based on the following two reduced form models,

Model 1. $x_i = 2z_{i1} + 0.75z_{i2} + 1.5z_{i3} + z_{i4} + \xi_i$;

Model 2. $x_i = 2z_{i1}^2 + 0.75z_{i2} + 1.5z_{i3}^2 + 3\sin(\pi z_{i4}) + \xi_i$,

where $\mathbf{z}_i = (z_{i1}, z_{i2}, \ldots, z_{ip})^{\mathrm{T}}$ is generated from a multivariate normal distribution $N(0, \Sigma)$, $\Sigma = (\rho_{j_1j_2})_{p\times p}$ with $\rho_{j_1j_2} = 0.5^{|j_1-j_2|}$, $j_1, j_2 = 1, \ldots, p$, and for each $i = 1, \ldots, n$. We generate the error terms in both the structural model and reduced form models by

$$(\varepsilon_i, \xi_i) \overset{\text{iid}}{\sim} N(0, \Sigma_{\varepsilon\xi}), \text{ with } \Sigma_{\varepsilon\xi} = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}.$$

We choose the sample size $n = 100, 200$ and the number of potential instruments $p = 100$. Note that there are only four valid instruments for the endogenous variable.

To assess the finite sample performance of each estimation method, we run each simulation $R = 1000$ times and computed the average of the estimated biases (denoted by "Bias"), $R^{-1}\sum_{r=1}^R(\widehat{\beta}_r - \beta_0)$, with its empirical standard deviation and the estimated mean squared errors (denoted by "MSE"), $R^{-1}\sum_{r=1}^R(\widehat{\beta}_r - \beta_0)^2$, where $\widehat{\beta}_r$ denotes an estimator of β_0 in the rth experiment.

All simulation studies are conducted using the statistical software R. In particular, the function *bs(z,degree,knots)* in the R package *splines* is used to obtain the B-splines series for each z. We vary in *degree* from 1 to 5 and choose knots = NULL or a vector of 25%, 50%, and 75% quantiles for each z, and then determine the optimal *degree* with/without knots by minimizing BIC values. It is worth noting that this data-driven procedure allows us to choose the optimal degree instead of arbitrarily setting a number in practice. It will be shown that, when the true reduced form model is indeed linear, the optimal degree is always chosen as 1 without knots; thus, the NAIVE automatically becomes the 2SLS estimator. Thus, our proposed method nests the traditional linear IV approach as a special case. We use the function *grpreg(z,x,group,penalty="grLasso")* in the R package *grpreg* (Breheny 2014) to estimate the additive reduced form model with the adaptive group Lasso.

In the first simulation, we assume it is known as an oracle that $z_{i1}, z_{i2}, z_{i3}, z_{i4}$ are the four truly valid instruments. This simulation is designed to check the influence of the nonparametric form of the instrumental variables on the estimation of the true effect β_0 given the perfect information on the strength of instruments. We consider three estimators of β_0: OLS, 2SLS, and NAIVE without variable selection procedure. Table 1 summarizes the averages of biases with the standard deviations in the parentheses and MSE values. Figure 1 shows the boxplots of biases for three estimators. It is shown that the OLS estimators are always biased and have the largest MSE due to the endogeneity issue. When the reduced form model is linear (Model 1), the 2SLS estimator solves the endogeneity problem well with small biases and MSE. However, when the relationship between the endogenous variable and the instruments is nonlinear (Model 2), its performance substantially deteriorates. On the other hand, the NAIVE performs the best in both linear and nonlinear settings. In particular, for the linear reduced form (Model 1), the NAIVE automatically chooses the degree to be 1 without knots and thus has the same performance as the 2SLS estimator.

In the second simulation, we suppose that there are $p = 100$ potential instruments but the researcher does not know which one is valid. Three estimators of β_0 are obtained: OLS, 2SLS based on the selected instruments by the Lasso (denoted by "2SLS-L"), and NAIVE with the adaptive group Lasso ("NAIVE"). The simulation results are summarized in Table 2 and Figure 2. As we observed in the first simulation, the NAIVE with the adaptive group Lasso almost agrees with the 2SLS estimator with the Lasso for Model 1. For the nonlinear

Table 1. The results of the first simulation with known truly valid instruments

		Model 1		Model 2	
n	Method	Bias	MSE	Bias	MSE
100	OLS	0.0527(0.0254)	0.0034	0.0418(0.0245)	0.0024
	2SLS	0.0016(0.0271)	0.0007	0.0149(0.1070)	0.0117
	NAIVE	0.0016(0.0271)	0.0007	0.0105(0.0244)	0.0007
200	OLS	0.0518(0.0175)	0.0030	0.0422(0.0166)	0.0021
	2SLS	−0.0001(0.0187)	0.0004	0.0135(0.0818)	0.0069
	NAIVE	−0.0001(0.0187)	0.0004	0.0059(0.0166)	0.0003

Table 2. The results of the second simulation in high-dimensional setting

		Model 1		Model 2	
n	Method	Bias	MSE	Bias	MSE
100	OLS	0.0532(0.0252)	0.0035	0.0429(0.0241)	0.0024
	2SLS-L	0.0193(0.0262)	0.0011	0.0406(0.0363)	0.0030
	NAIVE	0.0193(0.0262)	0.0011	0.0314(0.0256)	0.0016
200	OLS	0.0526(0.0182)	0.0031	0.0422(0.0172)	0.0021
	2SLS-L	0.0123(0.0186)	0.0005	0.0358(0.0346)	0.0025
	NAIVE	0.0122(0.0186)	0.0005	0.0175(0.0186)	0.0006

reduced form models, the NAIVE with the adaptive group Lasso has an excellent performance by achieving smaller biases and MSE. As n increases from 100 to 200, the bias and MSE of the NAIVE decrease substantially, which further confirms the consistency of the NAIVE in Theorem 3.1. Hence, we conclude that the proposed method is reliable and useful in practice, especially when the relationship between the endogenous variable and many instruments is unknown.

5. APPLICATIONS TO TRADE AND ECONOMIC GROWTH

Global economic integration and reducing international trade barriers are often hot topics of various economic summit forums as well as the focus of related literature (Rodrik 2000; Rose and van Wincoop 2001). The validity of those debating points and economic studies rely heavily on the causal relationship between trade and growth. In economics theory, international trade is believed to cause growth in countries dwelling at the technological frontier by using economies of scale as well as improving resource allocation efficiency. However, there might be two opposite effects (pro-growth and anti-growth) for countries behind the technological frontier (Grossman and Helpman 1991). On the one hand, trade may hamper the long run economic development of natural resource-abundant countries if that country specializes in primary products or raw materials such as minerals, crude oil, etc. (Matsuyama 1992), which may lead to the unfortunate "resource curse" (Sachs and Warner 2001). There are also concerns about the dynamic economies of scale such that infant industry protection are common in many developing countries. On the other hand, trade has positive effect on growth of the countries that are well behind the technological frontier and use trade for technology upgrade through spillover effects. The effect of trade on growth is a very important research topic in both theoretical and empirical economics, which have strong effect on trade policies. Accompanying the rapid growth and industrialization of developing countries, such as Brazil, China, India, Mexico, etc. since the 1980s and the recent surging development of African countries, such as Côte d'Ivoire, Ghana, Rwanda, etc. in the 2000s, many empirical studies have analyzed the important role of international trade in catch-up growth of those countries (Krueger 1990; Edwards 1997; Acemoglu and Ventura 2002; Dollar and Kraay 2004). Several influential studies have drawn quite different conclusions regarding trade and growth using the data from different decades (Sala-i-Martin 1997; Frankel and Romer 1999). The question of the relationship between trade and growth is still largely open in today's world.

In this section, we illustrate the use of NAIVE by revisiting the classic question of trade and growth. We explicitly investigate for the first time the role of trade in economic growth of 150 countries using nonparametric reduced form equations with instruments selection. One important issue in the empirical study of trade and growth is the endogeneity of trade variable due to the common driving forces that cause both trade and growth. Frankel and Romer (1999) (FR99 henceforth) showed that trade activities correlate positively with growth rate using cross-sectional data from 150 countries and economies. In

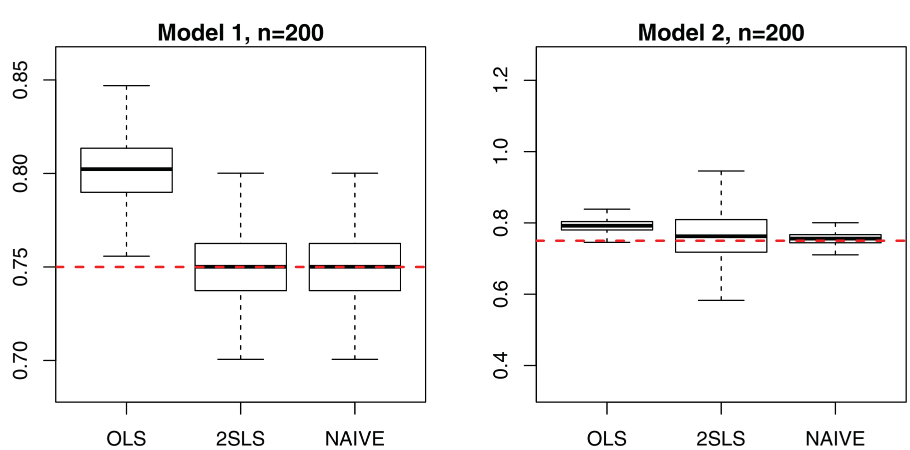


Figure 1. Boxplots of biases for the first simulation with known truly valid instruments.

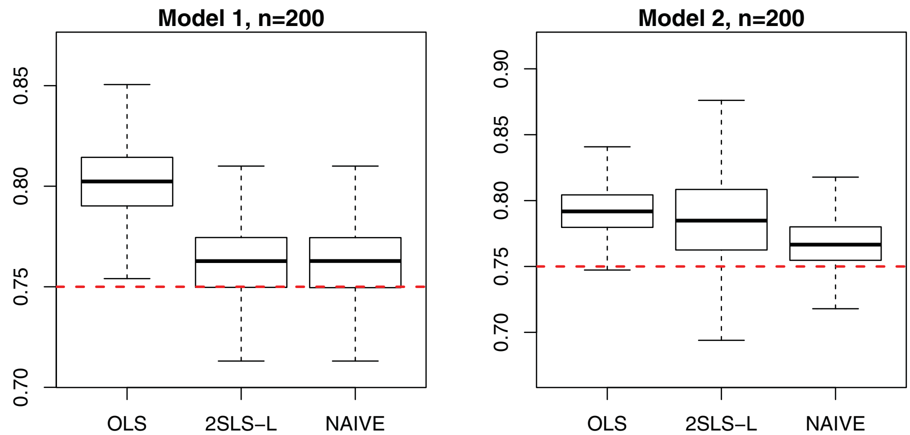


Figure 2. Boxplots of biases for the second simulation in high-dimensional setting.

FR99, the gravity model of trade (Tinbergen 1962; Anderson 1979) was applied to circumvent the endogeneity problem of trade using instrumental variable method. The gravity theory of trade states that the size of the countries (population, area, etc.) and the distance between them (up to a gravity parameter) determine the trade volume between two countries. Specifically, the instrument is the proxy variable for trade, which is constructed using geographical variables such as the country size, common border, and bilateral distance of two countries. The econometric model based on the gravity theory of trade is strongly supported by the data in empirical analysis (Disdier and Head 2008).

Following FR99, the structural equation we consider here is

$$lnY_i = \alpha + \boldsymbol{\beta} T_i + \boldsymbol{\gamma} S_i + \varepsilon_i$$

where Y_i is GDP per worker in country i, T_i is the share of international trade to GDP, S_i is the size of a country, and ε_i is the unobserved random disturbances, for $i = 1, 2, \ldots, 150$.

To find a valid instrumental variable for trade, we need to search for the variables that satisfy the two conditions. First, instruments have to be strictly exogenous to the structural equation of economic growth, and second, they have to be determinants of trade. In other words, to fit the empirical study, the instruments should affect growth only indirectly through trade. Under this logic and the gravity model of Tinbergen (1962), FR99 constructed the instruments using the distance between two countries. The validity of using geographical variables as instruments is justified by the following reasons. First, the geographical variables, such as distance between two countries, affect the convenience of trade through the channel of transportation costs among others. Second, the size and distance variables are fixed in the dataset for each country; thus, they are exogenous to the structural equation.

The bilateral trade reduced form equation, which FR99 employed to construct the instrumental variable, is a linear equation

$$T_{ik} = \boldsymbol{\theta} X_{ik} + \eta_{ik},$$

where the T_{ik} is the log of bilateral trade share of country i with country k ($k = 1, \ldots n, k \neq i$) and X_{ik} is a vector of instruments, which includes the distance between two countries, dummy variables for landlocked countries, common border between two countries and the interaction terms, and two included exogenous variables representing country size: population and area (Land). The instrumental variable (called proxy for trade in FR99) is the sum (over k) of predicted trade shares for country i.

We echo the two major problems commonly associated with finding instruments in empirical study using the example of trade. First, we usually do not know the quality of the instruments since there are many geographic variables; some or all of them might be irrelevant to the trade. The inclusion of irrelevant instruments distorts the nominal confidence interval of β. Second, finding some candidate instruments, in this example the geographic variables, does not ensure that the linear reduced form model would fully comprehend the nonlinearity nature of some instruments to endogenous variable. These problems will be addressed by our NAIVE method as discussed in previous sections.

In the following, we extend the cross-sectional study of FR99 by considering more potential instruments. Since we first want to include as many exogenous instruments as possible in the reduced form model, it is very likely that we would end up including some instruments that are not the determinants of trade. Through this real example, we show that our method could provide a guideline for empirical researchers on how to select instruments in nonparametric reduced form equation. Besides the three original instruments, that is, proxy for trade, total population, and total land area (included exogenous variables), we also include total water area, coastline, the arable land as percentage of total land, land boundaries, forest area as percentage of land area, the number of official and other commonly used languages in a country, and the interaction terms of constructed trade proxy with these variables (in total 15 instruments). All newly considered instruments are geographical variables, which are fixed in the dataset (hence exogenous), and they can only affect growth through the channel of trade.

The reduced form model we consider is

$$T_i = \mu + \sum_{j=1}^{15} f_j(z_{ij}) + \xi_i,$$

where $f_j(\cdot)$ is the jth unknown smooth univariate functions and z_{ij} is the ith observed value of the aforementioned jth instrument, $j = 1, 2, \ldots, 15$. Notice that we treat the interaction terms as a distinct instrument, and we allow the order of spline functions to differ for each *j*.

5.1 Data Description

We present the summary statistics of the main data in Table 3. For the summary statistics of bilateral distances and border, landlocked, and other variables used to construct the proxy for trade, we refer to the excellent presentation in FR99. To compare with the original study of FR99, we use the same data in this article and combine it with new geographical instrumental variables assessed in the same year (1985) as FR99. We would be able to include the most up-to-date data and search for new empirical support but here we will leave the new data analysis as future works that would use our method.

The aforementioned instruments could be potentially useful in the reduced form regression. However, it remains unknown which instrument is truly useful and the true functional form is unknown. Therefore, we apply the adaptive group Lasso to select instruments in the nonparametric additive reduced form model.

Table 3. Summary statistics

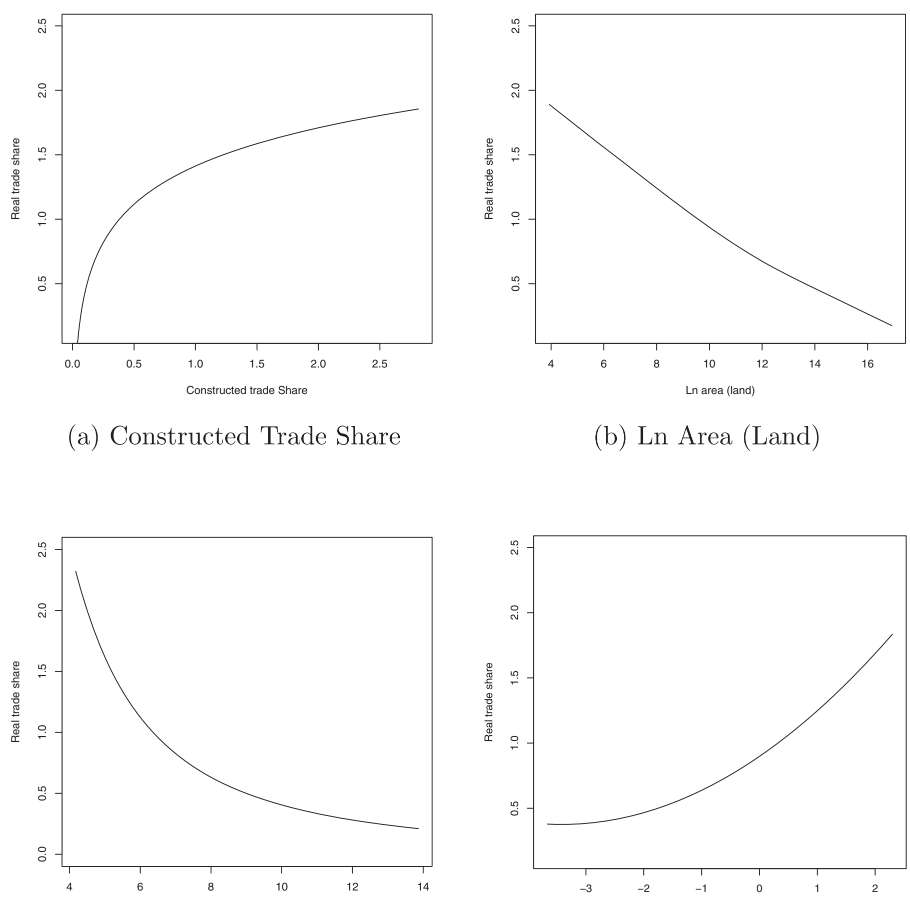
	Mean	sd	Median	Min	Max	Sample size
Ln GDP	8.81	1.04	8.87	6.56	10.55	150
Trade	0.73	0.46	0.63	0.13	3.18	150
Ln Population	8.61	1.93	8.78	4.17	13.87	150
Ln Area (Land)	11.71	2.42	12.19	3.93	16.92	150
Area (Water)	3.9E4	2.1E5	2.3E3	0	2.3E6	150
Coastline	4.6E3	1.8E4	654	0	2.0E5	150
Land Boundaries	2.9E3	3.7E3	2.0E3	0	2.2E4	150
% Forest	0.30	0.23	0.29	0	0.90	150
% Arable Land	0.14	0.13	0.10	0.01	0.58	150
Languages	3.79	6.44	2	1	52	150

NOTE: Water area, coastline, and land boundaries are measured in square kilometers and kilometers, respectively. Source: FR99, the World Bank, and CIA world Factbook.

5.2 Empirical Results

In this study, we focus on investigating the relationship of trade with growth. For the other results, such as the income distribution, the readers can refer to the original paper of FR99 and other excellent aforementioned papers.

Using the adaptive group Lasso and the BIC selection, the selected instruments include the proxy for trade (the original



(a) Constructed Trade Share (b) Ln Area (Land)

(c) Ln Population (d) Ln Const. Trade Share × Languages

Figure 3. Plots of the endogenous variable (real trade share) against the selected four instrumental variables.

Table 4. Estimation results for the trade and income data. Standard errors are reported in parentheses. Significance levels 0.1, 0.05, and 0.01 are noted by *, **, and *** respectively. Intercept significance levels are not reported.

	OLS	2SLS	2SLS-L	NAIVE
Constant	7.40	4.96	5.29	3.95
	(0.66)	(2.20)	(1.86)	(1.58)
Trade Share	0.85***	1.97**	1.71	**2.35***
	(0.25)	(0.99)	(0.79)	**(0.62)**
Ln Population	0.12**	0.19**	0.21	0.24*
	(0.06)	(0.09)	(0.09)	(0.14)
Ln Area	−0.01	0.09	0.04	0.08
	(0.06)	(0.10)	(0.07)	(0.08)
Sample Size	150	150	150	150

instrument in FR99), area of land, total population, and the interaction term of proxy for trade and number of languages. For the constructed trade share, degree 3 is selected. Land area has knots equal to 0 and hence is a linear fit. Population and the interaction terms of trade share and languages have degree 3. All the nonlinear fits use the quantiles as knots sequence. The fitted functions of selected instruments are plotted in Figure 3. From Figure 3, we see that proxy for trade and population instruments are likely to have nonlinear relationship with real trade share. Land size is more likely to have a linear functional form. From Panel (a) of Figure 3, we see that real trade share is in general increasing with constructed trade. The positive effect is also shown under the interaction term with languages as shown in Panel (d), and more languages correlate with higher openness to trade. The interaction term is selected to support the effects of language, which positively affects trade, given all other geographic characteristics of two countries are equal. Generally speaking, the constructed trade share has a positive relationship (but shrinking in magnitude) with the constructed trade share. Hence, we suspect that the proxy variable of trade is not a perfect proxy especially for international trade-oriented economies, such as Hong Kong, Luxemburg (with real trade share being 210%, 212% of total GDP, respectively), among others. From panel (b), the land area variable is negatively related with trade share. From panel (c), the log of population variable has dwindling effects on real trade share. Both represent a result of large countries that are less active in international trade (in the perspective of GDP contribution) during the middle of 1980s. Holding other variables constant, the domestic market, instead of global market, is the focus for those economies. According to those findings, we believe that the NAIVE method proposed in the article is necessary to analyze the data to get a more accurate estimation of the trade effects on growth.

Table 4 shows the regression results. The first two columns are originally from FR99. The OLS estimator has severe bias and is inconsistent because of the endogeneity issue. The second column is the 2SLS estimator using constructed trade proxy as instrument. The third column ("2SLS-L") represents the result of using the Lasso to select all available instruments. Subsequently, only the selected instruments are used to obtain the 2SLS estimator. The fourth column ("NAIVE") shows the proposed method with the adaptive group Lasso. The t statistics value for the NAIVE on trade is 3.79, compared to 2.16 for the linear IV regression with the Lasso and 1.98 for FR99, where only the constructed trade proxy as instrument was used. The NAIVE method provides more significant results regarding trade on growth. Therefore, we provide the stronger evidence to show that trade is positively correlated with economic growth in the middle of 1980s.

6. CONCLUSION

In this article, we consider the general nonlinear reduced form equation of IV regression using many instrumental variables, the dimensionality of which could be larger than the sample size. The proposed NAIVE with the adaptive group Lasso is root-n consistent, asymptotically normal, and efficient. Numerical studies have shown that the NAIVE has less bias and is more efficient compared to the 2SLS estimator in the nonlinear reduced form. If the reduced form is indeed linear, the proposed NAIVE adaptively becomes the 2SLS estimator. Thus, our proposed method nests the traditional linear IV approach as a special case. Furthermore, the implementation of our method is computationally efficient and easy to apply in practice using R packages. The STATA package *naivereg* is also available upon request. In the empirical study where we revisit the trade and growth question, our findings support the stronger positive effects of international trade on economic growth. Therefore, we suggest our NAIVE method with the adaptive group Lasso in empirical studies, which encounter many instruments without knowing their strength and their functional forms in the reduced form equation. Certain research topics along with this work remain open for future study. First, we may consider a nonparametric study of both the reduced form and structural form equations and nonseparable models. Second, the systematic study of the high-dimensional reduced form equation with instruments interactions could be another topic.

APPENDIX

Proof of Theorem 3.1. The NAIVE with the adaptive group Lasso is

$$\widehat{\boldsymbol{\beta}} = \left(\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{x}}_i\mathbf{x}_i^{\mathrm{T}}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{x}}_i y_i\right).$$

Note that $y_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \varepsilon_i$, thus,

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= \left(\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{x}}_i\mathbf{x}_i^{\mathrm{T}}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{x}}_i\mathbf{x}_i^{\mathrm{T}}\right)\boldsymbol{\beta}_0 \\ &\quad + \left(\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{x}}_i\mathbf{x}_i^{\mathrm{T}}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{x}}_i\varepsilon_i\right) \\ &= \boldsymbol{\beta}_0 + \left(\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{x}}_i\mathbf{x}_i^{\mathrm{T}}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{x}}_i\varepsilon_i\right). \qquad \text{(A.1)}\end{aligned}$$

Note that $D_{i\ell} = E(x_{i\ell}|\mathbf{z}_i) = \mu_\ell + \sum_{j\in A_\ell} f_{j\ell}(z_{ij})$ is the conditional expectation of ℓth endogenous variable given by the instrumental variables, and $\widehat{x}_{i\ell} = \widehat{\mu}_\ell + \sum_{j\in\widehat{A}_\ell}\widehat{f}_{nj\ell}(z_{ij})$. We denote that $\mathbf{D}_i = (D_{i1},\ldots,D_{id_e}, x_{id_e+1},\ldots,x_{id})^{\mathrm{T}}$, and

$\widehat{\mathbf{x}}_i = (\widehat{x}_{i1}, \ldots, \widehat{x}_{id_e}, x_{id_e+1}, \ldots, x_{id})^{\mathrm{T}}$. Thus,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n \|\mathbf{D}_i - \widehat{\mathbf{x}}_i\| &= \frac{1}{n}\sum_{i=1}^n \sqrt{\sum_{l=1}^{d_e} |D_{i\ell} - \widehat{x}_{i\ell}|^2} \\ &\le \sqrt{d_e} \max_{1\le\ell\le d_e} \frac{1}{n}\sum_{i=1}^n |D_{i\ell} - \widehat{x}_{i\ell}| \\ &= \sqrt{d_e} \max_{1\le\ell\le d_e} \frac{1}{n}\sum_{i=1}^n \Bigg|\mu_\ell + \sum_{j\in\mathcal{A}_\ell} f_{j\ell}(z_{ij}) \\ &\quad -\widehat{\mu}_\ell - \sum_{j\in\widehat{\mathcal{A}}_\ell} \widehat{f}_{nj\ell}(z_{ij})\Bigg| \\ &\le \sqrt{d_e} \max_{1\le\ell\le d_e} \frac{1}{n}\sum_{i=1}^n \Bigg\{|\widehat{\mu}_\ell - \mu_\ell| + \Bigg|\sum_{j\in\widehat{\mathcal{A}}_\ell} \widehat{f}_{nj\ell}(z_{ij}) \\ &\quad - \sum_{j\in\mathcal{A}_\ell} f_{j\ell}(z_{ij})\Bigg|\Bigg\} \\ &=: \sqrt{d_e} \max_{1\le\ell\le d_e} \frac{1}{n}\sum_{i=1}^n (S_{1\ell} + S_{2\ell}).\end{aligned}$$

By the central limit theorem, $S_{1\ell} = O_p(1/\sqrt{n}) = o_p(1)$. For any $\delta > 0$, we have

$$\begin{aligned}P(|S_{2\ell}| > \delta) &\le P\left(\left|\sum_{j\in\widehat{\mathcal{A}}_\ell\bigcap\mathcal{A}_\ell} [\widehat{f}_{nj\ell}(z_{ij}) - f_{j\ell}(z_{ij})]\right| > \delta/3\right) \\ &\quad + P\left(\left|\sum_{j\in\widehat{\mathcal{A}}_\ell\bigcap\mathcal{A}_\ell^c} \widehat{f}_{nj\ell}(z_{ij})\right| > \delta/3\right) \\ &\quad + P\left(\left|\sum_{j\in\widehat{\mathcal{A}}_\ell^c\bigcap\mathcal{A}_\ell} f_{j\ell}(z_{ij})\right| > \delta/3\right) \\ &\le P\left(\sqrt{\sum_{j\in\mathcal{A}_\ell} \left|\widehat{f}_{nj\ell}(z_{ij}) - f_{j\ell}(z_{ij})\right|^2} > \delta/3\right) \\ &\quad + P\left(\left|\sum_{j\in\widehat{\mathcal{A}}_\ell\bigcap\mathcal{A}_\ell^c} \widehat{f}_{nj\ell}(z_{ij})\right| > \delta/3\right) \\ &\quad + P\left(\left|\sum_{j\in\widehat{\mathcal{A}}_\ell^c\bigcap\mathcal{A}_\ell} f_{j\ell}(z_{ij})\right| > \delta/3\right) \\ &\to 0,\end{aligned}$$

where the first term $\to 0$ because $\sum_{j\in\mathcal{A}_\ell} |\widehat{f}_{nj\ell}(z_{ij}) - f_{j\ell}(z_{ij})|^2 = O(n^{-2s/(2s+1)}) = o_p(1)$ and the last two terms $\to 0$ because of the selection consistency of the adaptive group Lasso, that is, $P(\widehat{\mathcal{A}}_\ell = \mathcal{A}_\ell) \to 1$. Thus, $S_{2\ell} = o_p(1)$. Since d_e is fixed, we have $n^{-1}\sum_{i=1}^n \|\mathbf{D}_i - \widehat{\mathbf{x}}_i\| = o_p(1)$.

Since $E(x_{i\ell}^2)$ is bounded, $n^{-1}\sum_{i=1}^n x_{i\ell}^2 = E(x_{i\ell}^2) + O_p(1/\sqrt{n}) = O_p(1)$. Then, we have $n^{-1}\sum_{i=1}^n \|\mathbf{x}_i\|^2 = \sum_{\ell=1}^d n^{-1}\sum_{i=1}^n x_{i\ell}^2 = O_p(1)$ since d is a finite number. It implies that

$$\begin{aligned}\left\|\frac{1}{n}\sum_{i=1}^n \widehat{\mathbf{x}}_i\mathbf{x}_i^{\mathrm{T}} - \frac{1}{n}\sum_{i=1}^n \mathbf{D}_i\mathbf{x}_i^{\mathrm{T}}\right\| &\le \left\|\frac{1}{n}\sum_{i=1}^n (\widehat{\mathbf{x}}_i - \mathbf{D}_i)\mathbf{x}_i^{\mathrm{T}}\right\| \\ &\le \sqrt{\frac{1}{n}\sum_{i=1}^n \|\mathbf{x}_i\|^2 \frac{1}{n}\sum_{i=1}^n \|\widehat{\mathbf{x}}_i - \mathbf{D}_i\|^2} = o_p(1). \qquad \text{(A.2)}\end{aligned}$$

Next, we consider

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^n (\widehat{x}_{i\ell} - D_{i\ell})\varepsilon_i &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \left[\left(\widehat{\mu}_\ell + \sum_{j\in\widehat{\mathcal{A}}_\ell} \widehat{f}_{nj\ell}(z_{ij})\right)\right. \\ &\quad \left. - \left(\mu_\ell + \sum_{j\in\mathcal{A}_\ell} f_{j\ell}(z_{ij})\right)\right]\varepsilon_i \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n (\widehat{\mu}_\ell - \mu_\ell)\,\varepsilon_i \\ &\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\sum_{j\in\widehat{\mathcal{A}}_\ell} \widehat{f}_{nj\ell}(z_{ij}) - \sum_{j\in\mathcal{A}_\ell} f_{nj\ell}(z_{ij})\right)\varepsilon_i \\ &\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\sum_{j\in\mathcal{A}_\ell} f_{nj\ell}(z_{ij}) - \sum_{j\in\mathcal{A}_\ell} f_{j\ell}(z_{ij})\right)\varepsilon_i \\ &=: T_{1\ell} + T_{2\ell} + T_{3\ell},\end{aligned}$$

where $f_{nj\ell}(z_{ij}) = \sum_{k=1}^{m_n} \gamma_{jk}\psi_k(z_{ij})$ is the linear combination of normalized B-spline series for $f_{j\ell}(z_{ij})$. We first note that $T_{1\ell} = \frac{1}{\sqrt{n}}\sum_{i=1}^n (\widehat{\mu}_\ell - \mu_\ell)\,\varepsilon_i = (\widehat{\mu}_\ell - \mu_\ell)\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i = O_p(1/\sqrt{n})O_p(1) = o_p(1)$ due to the central limit theorem. Then, we deal with $T_{3\ell}$. Note that z_{ij} as the instrumental variable is uncorrelated with ε_i, so are $f_{nj\ell}(z_{ij})$ and $f_{j\ell}(z_{ij})$. Thus, we have

$$\begin{aligned}E(T_{3\ell}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n E\left[\left(\sum_{j\in\mathcal{A}_\ell} f_{nj\ell}(z_{ij}) - \sum_{j\in\mathcal{A}_\ell} f_{j\ell}(z_{ij})\right)\right. \\ &\quad \left.\times E\big(\varepsilon_i \big| z_{ij}, j\in\mathcal{A}_\ell\big)\right] = 0\end{aligned}$$

$$\begin{aligned}\mathrm{Var}(T_{3\ell}) &= \frac{1}{n}\sum_{i=1}^n Var\left\{E\left[\left(\sum_{j\in\mathcal{A}_\ell} f_{nj\ell}(z_{ij}) - \sum_{j\in\mathcal{A}_\ell} f_{j\ell}(z_{ij})\right)\right.\right. \\ &\quad \left.\left.\times\ \varepsilon_i \big| z_{ij}, j\in\mathcal{A}_\ell\right]\right\} \\ &\quad + \frac{1}{n}\sum_{i=1}^n E\left\{Var\left[\left(\sum_{j\in\mathcal{A}_\ell} f_{nj\ell}(z_{ij}) - \sum_{j\in\mathcal{A}_\ell} f_{j\ell}(z_{ij})\right)\right.\right. \\ &\quad \left.\left.\times\ \varepsilon_i \big| z_{ij}, j\in\mathcal{A}_\ell\right]\right\} \\ &= \sigma^2 E\left(\sum_{j\in\mathcal{A}_\ell} f_{nj\ell}(z_{ij}) - \sum_{j\in\mathcal{A}_\ell} f_{j\ell}(z_{ij})\right)^2 \\ &\le \sigma^2 E\left(q_\ell \sum_{j\in\mathcal{A}_\ell} [f_{nj\ell}(z_{ij}) - f_{j\ell}(z_{ij})]^2\right) = o_p(1),\end{aligned}$$

because q_ℓ is fixed and $[f_{nj\ell}(z_{ij}) - f_{j\ell}(z_{ij})]^2 = O_p(m_n^{-2s}) = o_p(1)$, which was proved in lemma 1 of Huang, Horowitz, and Wei (2010). Thus, by the Chebyshev's inequality, we have $T_{3\ell} = o_p(1)$.

Then, to deal with $T_{2\ell}$, we follow the idea of the proof of theorem 4 in Belloni et al. (2012) using moderate deviation inequality for self-normalized sums. We first present the following lemma—lemma 5 in Belloni et al. (2012), which was also based on theorem 7.4 in de la Pena, Lai, and Shao (2009). □

Lemma A.1. Let $X_1, \ldots, X_n$ be the triangular array of iid zero-mean random variables. Suppose that $M_n = (EX_1^2)^{1/2}/(E|X_1|^3)^{1/3} > 0$ and that for some $b_n \to \infty$ slowly, $n^{1/6}M_n/b_n \geq 1$. Then uniformly on $0 \leq x \leq n^{1/6}M_n/b_n - 1$, we have

$$\left|\frac{P(|S_n/V_n| \geq x)}{2[1-\Phi(x)]} - 1\right| \leq \frac{A}{b_n^3} \to 0,$$

where $S_n = \sum_{i=1}^n X_i$, $V_n^2 = \sum_{i=1}^n X_i^2$, $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution and A is a positive constant.

Now, we consider the term $T_{2\ell}$,

$$\begin{aligned}|T_{2\ell}| &= \left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\sum_{j\in\widehat{\mathcal{A}}_\ell}\sum_{k=1}^{m_n}\widehat{\gamma}_{jk\ell}\psi_k(z_{ij}) - \sum_{j\in\mathcal{A}_\ell}\sum_{k=1}^{m_n}\gamma_{jk\ell}\psi_k(z_{ij})\right)\varepsilon_i\right|\\
&= \left|\sum_{j\in\widehat{\mathcal{A}}_\ell\cup\mathcal{A}_\ell}\sum_{k=1}^{m_n}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\psi_k(z_{ij})\varepsilon_i\right)(\widehat{\gamma}_{jk\ell} - \gamma_{jk\ell})\right|\\
&\leq \max_{j\in\widehat{\mathcal{A}}_\ell\cup\mathcal{A}_\ell, 1\leq k\leq m_n}\left|\frac{\sum_{i=1}^n\psi_k(z_{ij})\varepsilon_i}{\sqrt{\sum_{i=1}^n\psi_k^2(z_{ij})\varepsilon_i^2}}\right|\\
&\quad\times \max_{j\in\widehat{\mathcal{A}}_\ell\cup\mathcal{A}_\ell, 1\leq k\leq m_n}\sqrt{\frac{1}{n}\sum_{i=1}^n\psi_k^2(z_{ij})\varepsilon_i^2}\\
&\quad\times \sum_{j\in\widehat{\mathcal{A}}_\ell\cup\mathcal{A}_\ell}\sum_{k=1}^{m_n}|\widehat{\gamma}_{jk\ell} - \gamma_{jk\ell}|.\end{aligned}$$

Note that

$$\begin{aligned}&P\left(\max_{j\in\widehat{\mathcal{A}}_\ell\cup\mathcal{A}_\ell, 1\leq k\leq m_n}\left|\frac{\sum_{i=1}^n\psi_k(z_{ij})\varepsilon_i}{\sqrt{\sum_{i=1}^n\psi_k^2(z_{ij})\varepsilon_i^2}}\right| > \sqrt{2m_n/a}\right)\\
&\leq |\widehat{\mathcal{A}}_\ell\cup\mathcal{A}_\ell|m_n \max_{j\in\widehat{\mathcal{A}}_\ell\cup\mathcal{A}_\ell, 1\leq k\leq m_n} P\left(\left|\frac{\sum_{i=1}^n\psi_k(z_{ij})\varepsilon_i}{\sqrt{\sum_{i=1}^n\psi_k^2(z_{ij})\varepsilon_i^2}}\right| > \sqrt{2m_n/a}\right)\\
&\leq q_\ell(1+o(1))m_n 2\left[1-\Phi\left(\sqrt{2m_n/a}\right)\right](1+o(1))\\
&\leq q_\ell m_n\frac{2\exp\left[-\left(\sqrt{2m_n/a}\right)^2/2\right]}{\sqrt{2\pi}\sqrt{2m_n/a}}(1+o(1))\\
&= \frac{q_\ell\sqrt{m_n a}}{\sqrt{\pi}\exp(m_n/a)}(1+o(1)) \to 0,\end{aligned}$$

uniformly for all $0 < a \leq 1$, where the second inequality follows the above lemma on moderate deviation inequality for self-normalized sums and the last inequality follows the fact that $P(Z > z) \leq \exp(-z^2/2)/(z\sqrt{2\pi})$ for a standard normal random variable Z. Thus,

$$\max_{j\in\widehat{\mathcal{A}}_\ell\cup\mathcal{A}_\ell, 1\leq k\leq m_n}\left|\frac{\sum_{i=1}^n\psi_k(z_{ij})\varepsilon_i}{\sqrt{\sum_{i=1}^n\psi_k^2(z_{ij})\varepsilon_i^2}}\right| = O_p(\sqrt{m_n}) = O_p(n^{1/(4s+2)}). \tag{A.3}$$

Since the centered B-splines $|\psi_k(z_{ij})| \leq 2$ and $E(\varepsilon_i^2)$ is bounded,

$$\begin{aligned}&\max_{j\in\widehat{\mathcal{A}}_\ell\cup\mathcal{A}_\ell, 1\leq k\leq m_n}\sqrt{\frac{1}{n}\sum_{i=1}^n\psi_k^2(z_{ij})\varepsilon_i^2}\\
&\leq \max_{j\in\widehat{\mathcal{A}}_\ell\cup\mathcal{A}_\ell, 1\leq k\leq m_n} 2\sqrt{\frac{1}{n}\sum_{i=1}^n\varepsilon_i^2} = O_p(1).\end{aligned} \tag{A.4}$$

Because of the selection consistency and estimation consistency in Lemma 3.1,

$$\begin{aligned}\sum_{j\in\widehat{\mathcal{A}}_\ell\cup\mathcal{A}_\ell}\sum_{k=1}^{m_n}|\widehat{\gamma}_{jk\ell} - \gamma_{jk\ell}| &\leq \sqrt{q_\ell(1+o(1))m_n\sum_{j\in\widehat{\mathcal{A}}_\ell\cup\mathcal{A}_\ell}\sum_{k=1}^{m_n}(\widehat{\gamma}_{jk\ell} - \gamma_{jk\ell})^2}\\
&= \sqrt{q_\ell(1+o(1))m_n\left[O_p\left(n^{-(2s-1)/(2s+1)}\right)\right]}\\
&= O_p\left(n^{-(2s-2)/(4s+2)}\right).\end{aligned} \tag{A.5}$$

Thus, (A.3), (A.4), and (A.5) together imply that

$$\begin{aligned}|T_{2\ell}| &= O_p\left(n^{1/(4s+2)}\right)O_p(1)O_p\left(n^{-(2s-2)/(4s+2)}\right)\\
&= O_p\left(n^{-(2s-3)/(4s+2)}\right) = o_p(1),\end{aligned} \tag{A.6}$$

provided that $s > 1.5$. Therefore, we have $|T_{t\ell}| = o_p(1)$, $t = 1, 2, 3$, which further imply that

$$\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n(\widehat{x}_{i\ell} - D_{i\ell})\varepsilon_i\right\| = o_p(1). \tag{A.7}$$

Therefore, (A.1) together with (A.2) and (A.7) imply that

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) = \left(\frac{1}{n}\sum_{i=1}^n\mathbf{D}_i\mathbf{x}_i^{\mathrm{T}} + o_p(1)\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\mathbf{D}_i\varepsilon_i + o_p(1)\right).$$

Because that $n^{-1}\sum_{i=1}^n\mathbf{D}_i\mathbf{x}_i^{\mathrm{T}} \xrightarrow{p} E(\mathbf{D}_i\mathbf{x}_i^{\mathrm{T}}) = E(\mathbf{D}_i\mathbf{D}_i^{\mathrm{T}})$ by the Weak Law of Large Numbers, and $D_i\varepsilon_i$ are iid with mean zero and variance $\sigma^2E(\mathbf{D}_i\mathbf{D}_i^{\mathrm{T}})$, the Central Limit Theorem and Slutsky's Theorem together imply that

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \xrightarrow{d} N\left(0, \sigma^2\left\{E\left(\mathbf{D}_i\mathbf{D}_i^{\mathrm{T}}\right)\right\}^{-1}\right),$$

which completes the proof of Theorem 3.1. □

ACKNOWLEDGMENTS

Qingliang Fan's research was supported by National Natural Science Foundation of China (NNSFC) grant 71301134 and Natural Science Foundation of Fujian Province of China (No. 2016J01340). Wei Zhong's research was supported by NNSFC grants 11301435 and 11401497, and the Fundamental Research Funds for the Central Universities 20720140034. Both authors equally contributed to this article. We are grateful to Whitney Newey and Shakeeb Khan for their helpful comments. The authors also thank the editor, the associate editor, and two reviewers for their constructive comments, which have greatly improved the earlier version of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NNSFC.

[Received January 2015. Revised March 2016.]

REFERENCES

Acemoglu, D., and Ventura, J. (2002), "The World Income Distribution," *Quarterly Journal of Economics*, 117, 659–694. [393]

Amemiya, T. (1974), "The Non-Linear Two-Stage Least Squares Estimator," *Journal of Econometrics*, 2, 105–110. [388,389,392]

Anderson, J. A. (1979), "A Theoretical Foundation for the Gravity Equation," *American Economic Review*, 69, 106–116. [394]

Andrews, D., Moreira, M., and Stock, J. (2006), "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression," *Econometrica*, 74, 715–752. [388]

Angrist, J., and Krueger, A. (1991), "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics*, 106, 979–1014. [388]

Bai, J., and Ng, S. (2010), "Instrumental Variable Estimation in a Data Rich Environment," *Econometric Theory*, 26, 1577–1606. [389]

Behrman, J., Olivia, M., Soo, C., and Bravo, D. (2012), "How Financial Literacy Affects Household Wealth Accumulation," *American Economic Review*, 102, 300–304. [388]

Belloni, A., Chen, D., Chernozhukov V., and Hansen, C. (2012), "Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain," *Econometrica*, 80, 2369–2429. [389,392,398]

Berkowitz, D., Caner, M., and Fang, Y. (2012), "The Validity of Instruments Revisited," *Journal of Econometrics*, 166, 255–266. [388]

Bound, J., Jaeger, D. A., and Baker, R. (1995), "Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variables is Weak," *Journal of the American Statistical Association*, 90, 443–450. [388]

Breheny, P. (2014), R Package grpreg: Regularization Paths for Regression Models with Grouped Covariates, available at *http://cran.r-project.org/web/packages/grpreg/index.html* [392]

Candes, E., and Tao, T. (2007), "The Dantzig Selector: Statistical Estimation When p is Much Larger Than n," *The Annals of Statistics*, 35, 2313–2351. [389]

Caner, M., and Fan, Q. (2015), "Hybrid Generalized Empirical Likelihood Estimators: Instrument Selection With Adaptive Lasso," *Journal of Econometrics*, 187, 256–274. [389]

Carrasco, M. (2012), "A Regularization Approach to the Many Instruments Problem," *Journal of Econometrics*, 170, 383–398. [389]

Chamberlain, G. (1987), "Asymptotic Efficiency in Estimation With Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305–334. [389,392]

Chao, J., and Swanson, N. (2005), "Consistent Estimation With a Large Number of Weak Instruments," *Econometrica*, 73, 1673–1692. [388]

Chen, J., and Chen, Z. (2008), "Extended Bayesian Information Criteria for Model Selection With Large Model Spaces," *Biometrika*, 95, 759–771. [391]

de la Pena, V., Lai, T., and Shao, Q. (2009), *Self-Normalized Processes: Limit Theory and Statistical Applications*, New York: Springer-Verlag. [398]

Disdier, A., and Head, K. (2008), "The Puzzling Persistence of the Distance Effect on Bilateral Trade," *The Review of Economics and Statistics*, 90, 37–48. [394]

Dollar, D., and Kraay, A. (2004), "Trade, Growth and Poverty," *Economic Journal*, 114, 22–49. [393]

Donald, S. G., and Newey, W. (2001), "Choosing the Number of Instruments," *Econometrica*, 69, 1161–1191. [388]

Edwards, S. (1997), "Trade Policy, Growth, and Income Distribution," *American Economic Review*, 87, 205–210. [393]

Fan, J., Feng, Y., and Song, R. (2011), "Nonparametric Independence Screening in Sparse Ultra-High Dimensional Additive Models," *Journal of the American Statistical Association*, 106, 544–557. [391]

Fan, J., and Li, R. (2001), "Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties," *Journal of the American Statistical Association*, 96, 1348–1360. [389]

Fan, J., and Liao, Y. (2014), "Endogeneity in High Dimensions," *The Annals of Statistics*, 42, 872–917. [389]

Frankel, J., and Romer, D. (1999), "Does Trade Cause Growth?" *American Economic Review*, 89, 379–399. [393]

Grossman, G., and Helpman, E. (1991), *Innovation and Growth in the Global Economy*, Cambridge, MA: MIT Press. [393]

Hahn, J., and Hausman, J. (2002), "A New Specification Test for the Validity of Instrumental Variables," *Econometrica*, 70, 163–189. [388]

Hansen, C., Hausman, J., and Newey, W. (2008), "Estimation With Many Instrumental Variables," *Journal of Business and Economic Statistics*, 26, 398–422. [388]

Hausman, J., Newey, W., Woutersen, T., Chao, J., and Swanson, N. (2012), "Instrumental Variable Estimation with Heteroskedasticity and Many Instruments," *Quantitative Economics*, 3, 211–255. [388]

Huang, J., Horowitz, J., and Wei, F. (2010), "Variable Selection in Nonparametric Additive Models," *The Annals of Statistics*, 38, 2282–2313. [389,390,391,398]

Krueger, A. (1990), "Asian Trade and Growth Lessons," *American Economic Review*, 80, 108–112. [393]

Lin, W., Feng, R., and Li, H. (2015), "Regularization Methods for High-Dimensional Instrumental Variables Regression With an Application to Genetical Genomics," *Journal of the American Statistical Association*, 110, 270–288. [389]

Linton, O., Chen, R., Wang, N., and Härdle, W. (1997), "An Analysis of Transformations for Additive Nonparametric Regression," *Journal of the American Statistical Association*, 92, 1512–1521. [389]

Matsuyama, K. (1992), "Agricultural Productivity, Comparative Advantage, and Economic Growth," *Journal of Economic Theory*, 58, 317–334. [393]

Melitz, M. (2003), "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71, 1695–1725. [388]

Newey, W. (1990), "Efficient Instrumental Variables Estimation of Nonlinear Models," *Econometrica*, 58, 809–837. [388,389,390,392]

Newey, W., and Powell, J. (2003), "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578. [388]

Ozabaci, D., Henderson, D., and Su, L. (2014), "Additive Nonparametric Regression in the Presence of Endogenous Regressors," *Journal of Business and Economic Statistics*, 32, 555–575. [388]

Rodrik, D. (2000), "How Far Will International Economic Integration Go?" *The Journal of Economic Perspectives*, 14, 177–186. [393]

Rose, A., and van Wincoop, E. (2001), "National Money as a Barrier to International Trade: The Real Case for Currency Union," *American Economic Review*, 91, 386–390. [393]

Sachs, J., and Warner, A. (2001), "The Curse of Natural Resources," *European Economic Review*, 45, 827–838. [393]

Sala-i-Martin, X. (1997), "I Just Ran Two Million Regressions," *American Economic Review*, 87, 178–183. [393]

Schwartz, G. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461–464. [391]

Staiger, D., and Stock, J. (1997), "Instrumental Variables Regression With Weak Instruments," *Econometrica*, 65, 557–586. [388]

Stone, C. (1985), "Additive Regression and Other Nonparametric Models," *The Annals of Statistics*, 13, 689–705. [390]

Tibshirani, R. (1996), "Regression Shrinkage and Selection via Lasso," *Journal of the Royal Statistical Society,* Series B, 58, 267–288. [389]

Tinbergen, J. (1962), *Shaping the World Economy: Suggestions for an International Economic Policy*, New York: The Twentieth Century Fund. [394]

Wang, H., Li, R., and Tsai, C. (2007), "Tuning Parameter Selectors for the Smoothly Clipped Absolute Deviation Method," *Biometrika*, 94, 553–568. [391]

Yuan, M., and Lin, Y. (2006), "Model Selection and Estimation in Regression With Grouped Variables," *Journal of the Royal Statistical Society,* Series B, 68, 49–67. [389]

Zou, H. (2006), "The Adaptive Lasso and Its Oracle Properties," *Journal of the American Statistical Association*, 101, 1418–1429. [389]

IRTG 1792 Discussion Paper Series 2018

For a complete list of Discussion Papers published, please visit irtg1792.hu-berlin.de.

001 "Data Driven Value-at-Risk Forecasting using a SVR-GARCH-KDE Hybrid" by Marius Lux, Wolfgang Karl Härdle and Stefan Lessmann, January 2018.
002 "Nonparametric Variable Selection and Its Application to Additive Models" by Zheng-Hui Feng, Lu Lin, Ruo-Qing Zhu asnd Li-Xing Zhu, January 2018.
003 "Systemic Risk in Global Volatility Spillover Networks: Evidence from Option-implied Volatility Indices " by Zihui Yang and Yinggang Zhou, January 2018.
004 "Pricing Cryptocurrency options: the case of CRIX and Bitcoin" by Cathy YH Chen, Wolfgang Karl Härdle, Ai Jun Hou and Weining Wang, January 2018.
005 "Testing for bubbles in cryptocurrencies with time-varying volatility" by Christian M. Hafner, January 2018.
006 "A Note on Cryptocurrencies and Currency Competition" by Anna Almosova, January 2018.
007 "Knowing me, knowing you: inventor mobility and the formation of technology-oriented alliances" by Stefan Wagner and Martin C. Goossen, February 2018.
008 "A Monetary Model of Blockchain" by Anna Almosova, February 2018.
009 "Deregulated day-ahead electricity markets in Southeast Europe: Price forecasting and comparative structural analysis" by Antanina Hryshchuk, Stefan Lessmann, February 2018.
010 "How Sensitive are Tail-related Risk Measures in a Contamination Neighbourhood?" by Wolfgang Karl Härdle, Chengxiu Ling, February 2018.
011 "How to Measure a Performance of a Collaborative Research Centre" by Alona Zharova, Janine Tellinger-Rice, Wolfgang Karl Härdle, February 2018.
012 "Targeting customers for profit: An ensemble learning framework to support marketing decision making" by Stefan Lessmann, Kristof Coussement, Koen W. De Bock, Johannes Haupt, February 2018.
013 "Improving Crime Count Forecasts Using Twitter and Taxi Data" by Lara Vomfell, Wolfgang Karl Härdle, Stefan Lessmann, February 2018.
014 "Price Discovery on Bitcoin Markets" by Paolo Pagnottoni, Dirk G. Baur, Thomas Dimpfl, March 2018.
015 "Bitcoin is not the New Gold - A Comparison of Volatility, Correlation, and Portfolio Performance" by Tony Klein, Hien Pham Thu, Thomas Walther, March 2018.
016 "Time-varying Limit Order Book Networks" by Wolfgang Karl Härdle, Shi Chen, Chong Liang, Melanie Schienle, April 2018.
017 "Regularization Approach for NetworkModeling of German EnergyMarket" by Shi Chen, Wolfgang Karl Härdle, Brenda López Cabrera, May 2018.
018 "Adaptive Nonparametric Clustering" by Kirill Efimov, Larisa Adamyan, Vladimir Spokoiny, May 2018.
019 "Lasso, knockoff and Gaussian covariates: a comparison" by Laurie Davies, May 2018.

IRTG 1792, Spandauer Straße 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.

IRTG 1792 Discussion Paper Series 2018

For a complete list of Discussion Papers published, please visit irtg1792.hu-berlin.de.

020 "A Regime Shift Model with Nonparametric Switching Mechanism" by Haiqiang Chen, Yingxing Li, Ming Lin and Yanli Zhu, May 2018.
021 "LASSO-Driven Inference in Time and Space" by Victor Chernozhukov, Wolfgang K. Härdle, Chen Huang, Weining Wang, June 2018.
022 " Learning from Errors: The case of monetary and fiscal policy regimes" by Andreas Tryphonides, June 2018.
023 "Textual Sentiment, Option Characteristics, and Stock Return Predictability" by Cathy Yi-Hsuan Chen, Matthias R. Fengler, Wolfgang Karl Härdle, Yanchu Liu, June 2018.
024 "Bootstrap Confidence Sets For Spectral Projectors Of Sample Covariance" by A. Naumov, V. Spokoiny, V. Ulyanov, June 2018.
025 "Construction of Non-asymptotic Confidence Sets in 2 -Wasserstein Space" by Johannes Ebert, Vladimir Spokoiny, Alexandra Suvorikova, June 2018.
026 "Large ball probabilities, Gaussian comparison and anti-concentration" by Friedrich Götze, Alexey Naumov, Vladimir Spokoiny, Vladimir Ulyanov, June 2018.
027 "Bayesian inference for spectral projectors of covariance matrix" by Igor Silin, Vladimir Spokoiny, June 2018.
028 "Toolbox: Gaussian comparison on Eucledian balls" by Andzhey Koziuk, Vladimir Spokoiny, June 2018.
029 "Pointwise adaptation via stagewise aggregation of local estimates for multiclass classification" by Nikita Puchkin, Vladimir Spokoiny, June 2018.
030 "Gaussian Process Forecast with multidimensional distributional entries" by Francois Bachoc, Alexandra Suvorikova, Jean-Michel Loubes, Vladimir Spokoiny, June 2018.
031 "Instrumental variables regression" by Andzhey Koziuk, Vladimir Spokoiny, June 2018.
032 "Understanding Latent Group Structure of Cryptocurrencies Market: A Dynamic Network Perspective" by Li Guo, Yubo Tao and Wolfgang Karl Härdle, July 2018.
033 "Optimal contracts under competition when uncertainty from adverse selection and moral hazard are present" by Natalie Packham, August 2018.
034 "A factor-model approach for correlation scenarios and correlation stress-testing" by Natalie Packham and Fabian Woebbeking, August 2018.
035 "Correlation Under Stress In Normal Variance Mixture Models" by Michael Kalkbrener and Natalie Packham, August 2018.
036 "Model risk of contingent claims" by Nils Detering and Natalie Packham, August 2018.
037 "Default probabilities and default correlations under stress" by Natalie Packham, Michael Kalkbrener and Ludger Overbeck, August 2018.
038 "Tail-Risk Protection Trading Strategies" by Natalie Packham, Jochen Papenbrock, Peter Schwendner and Fabian Woebbeking, August 2018.

IRTG 1792, Spandauer Straße 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.

IRTG 1792 Discussion Paper Series 2018

For a complete list of Discussion Papers published, please visit irtg1792.hu-berlin.de.

039 "Penalized Adaptive Forecasting with Large Information Sets and Structural Changes" by Lenka Zbonakova, Xinjue Li and Wolfgang Karl Härdle, August 2018.

040 "Complete Convergence and Complete Moment Convergence for Maximal Weighted Sums of Extended Negatively Dependent Random Variables" by Ji Gao YAN, August 2018.

041 "On complete convergence in Marcinkiewicz-Zygmund type SLLN for random variables" by Anna Kuczmaszewska and Ji Gao YAN, August 2018.

042 "On Complete Convergence in Marcinkiewicz-Zygmund Type SLLN for END Random Variables and its Applications" by Ji Gao YAN, August 2018.

043 "Textual Sentiment and Sector specific reaction" by Elisabeth Bommes, Cathy Yi-Hsuan Chen and Wolfgang Karl Härdle, September 2018.

044 "Understanding Cryptocurrencies" by Wolfgang Karl Härdle, Campbell R. Harvey, Raphael C. G. Reule, September 2018.

045 "Predicative Ability of Similarity-based Futures Trading Strategies" by Hsin-Yu Chiu, Mi-Hsiu Chiang, Wei-Yu Kuo, September 2018.

046 "Forecasting the Term Structure of Option Implied Volatility: The Power of an Adaptive Method" by Ying Chen, Qian Han, Linlin Niu, September 2018.

047 "Inferences for a Partially Varying Coefficient Model With Endogenous Regressors" by Zongwu Cai, Ying Fang, Ming Lin, Jia Su, October 2018.

048 "A Regime Shift Model with Nonparametric Switching Mechanism" by Haiqiang Chen, Yingxing Li, Ming Lin, Yanli Zhu, October 2018.

049 "Strict Stationarity Testing and GLAD Estimation of Double Autoregressive Models" by Shaojun Guo, Dong Li, Muyi Li, October 2018.

050 "Variable selection and direction estimation for single-index models via DC-TGDR method" by Wei Zhong, Xi Liu, Shuangge Ma, October 2018.

051 "Property Investment and Rental Rate under Housing Price Uncertainty: A Real Options Approach" by Honglin Wang, Fan Yu, Yinggang Zhou, October 2018.

052 "Nonparametric Additive Instrumental Variable Estimator: A Group Shrinkage Estimation Perspective" by Qingliang Fan, Wei Zhong, October 2018.

IRTG 1792, Spandauer Straße 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.