UNIVERSITY OF SURREY

Discussion Papers in Economics

BOOTSTRAP INFERENCE FOR PENALIZED GMM ESTIMATORS WITH ORACLE PROPERTIES

By

Lorenzo Camponovo
(University of Surrey)

DP 06/18

School of Economics
University of Surrey
Guildford
Surrey GU2 7XH, UK
Telephone +44 (0)1483 689380
Facsimile +44 (0)1483 689548
Web www.econ.surrey.ac.uk

ISSN: 1749-5075

Bootstrap Inference for Penalized GMM Estimators with Oracle Properties

Lorenzo Camponovo[*]
University of Surrey

July 2017

Abstract

We study the validity of bootstrap methods in approximating the sampling distribution of penalized GMM estimators with oracle properties. More precisely, we focus on bridge estimators with L_q penalty for $0 < q < 1$, and adaptive lasso estimators. We show that the nonparametric bootstrap with recentered moment conditions provides a valid method for approximating the distribution of these estimators. Furthermore, using the bootstrap approach, we also propose a data-driven method for the selection of tuning parameters in the penalization terms. Monte Carlo simulations confirm the reliability and accuracy of the bootstrap procedure. The empirical coverages for the active variables implied by the nonparametric bootstrap are always very close to the nominal coverage probabilities.

JEL Classification: C12, C13, C52.
Keywords: Nonparametric Bootstrap, Penalized GMM Estimators, Oracle Properties.

[*]E-mail: l.camponovo@surrey.ac.uk. Address: Guildford, Surrey, GU2 7XH, UK.

1 Introduction

Since Hansen (1982), the generalized method of moments (GMM) has attained widespread applicability in various statistics and econometrics problems. The GMM provides a powerful tool for introducing statistical inference in several economic and financial models that are specified by some moment conditions; see, e.g., Hall (2005) for a review of the GMM. Important examples include instrumental variable regression models, where the GMM makes use of the orthogonality conditions to allow for efficient estimation in the presence of heteroskedasticity of unknown form.

In spite of a consistent (and efficient) estimate of the parameter of interest, GMM estimators do not perform variable selection. To overcome this problem, in the spirit of Frank and Friedman (1993), Tibshirani (1996), Fu (1998), Knight and Fu (2000), Fan and Li (2001) and Zou (2006), among others, recent research proposes penalized GMM estimators; see, e.g., Caner (2009) and Caner and Zhang (2014). By adding appropriate penalization terms in the GMM estimation criterion, penalized GMM estimators may achieve the so-called oracle properties. More precisely, they may simultaneously perform correct variable selection, and provide efficient estimates of the nonzero coefficients as if only the relevant variables had been included in the model.

In this paper, we study the validity of bootstrap methods in approximating the sampling distribution of penalized GMM estimators with oracle properties. More precisely, we focus on bridge estimators with L_q penalty for $0 < q < 1$, and adaptive lasso estimators. In particular, we show that the nonparametric bootstrap with recentered moments conditions provides a valid method for approximating the distribution of these estimators. Furthermore, using the bootstrap approach, we also propose a data-driven method for the selection of tuning parameters in the penalization terms. In our study, we mainly consider settings where the dimension of the parameter of interest d_θ and the number of moment conditions d_g are fixed, while the sample size n is large. However, in principle the nonparametric bootstrap can be applied also to settings where both d_θ and d_g may depend on the sample size n.

Recently, many authors have proposed inference procedures for penalized estimators, in particular in linear models. Some of these procedures rely on resampling methods, such as Chatterjee and Lahiri (2010, 2011, 2013), Minnier et al. (2011) and Camponovo (2015). Another class of inference methods considers a desparsification approach, which removes the bias introduced by shrinkage and constructs appropriate approximate inverses of the empirical Gram matrix; see, e.g., Zhang and Zhang (2014) and van de Geer et al. (2014). Lockhart et al. (2014) proposed a covariance statistic for testing the significance of predictors. Taylor et al. (2014) introduced a new class of test statistics for forward stepwise and least-angle regression that produces exact post-model-selection p-values. Belloni et al. (2014 a,b) defined inference methods for single coefficients that explicitly account for inevitable model selection errors in the asymptotic approximations. Our work is mostly related to that of Camponovo (2015), which focuses on the nonparametric bootstrap for lasso estimators in heteroskedastic linear regression models. Our results supplement previous findings by studying the validity of the nonparametric bootstrap for more general (nonlinear) penalized GMM estimators.

Finally, we study the accuracy of the nonparametric bootstrap through Monte Carlo simulations in instrumental variable regression models. The bootstrap approach ensures very accurate inference for the active variables. Indeed, the empirical coverages for the active variables implied by the nonparametric bootstrap are always very close to the nominal coverage probabilities. Furthermore, for the inactive variables the nonparametric bootstrap provides very short confidence intervals with empirical coverage converging to 1. Indeed, because of the variable selection property, the penalized GMM estimates of zero coefficients collapse to 0 asymptotically. Therefore, in this case the empirical coverage of confidence intervals should converge to 1 as n increases; see e.g., Minnier et al. (2011) for similar empirical findings.

The rest of the paper is organized as follows. In Section 2, we introduce the model and notation. In Section 3, we study the validity of bootstrap approximations for the sampling distribution of penalized GMM estimators with oracle properties. In Section 4, we propose a data-driven method for the selection of the tuning parameter. In Section 5, we present the Monte Carlo findings. Finally, Section 6 concludes.

2 Penalized GMM Estimators

In Section 2.1, we introduce the penalized GMM estimators, while in Section 2.2 we present their asymptotic properties.

2.1 Model and Notation

Let $(X_1, \dots, X_n)$ be an iid random sample from a probability distribution F. Consider the moment conditions $E[g(X_i, \theta_0)] = 0$, where $g(\cdot, \cdot)$ is a $\mathbb{R}^{d_g}$-valued function, θ_0 is the true value of the unknown parameter $\theta \in \Theta \subset \mathbb{R}^{d_\theta}$, and $d_g \geq d_\theta$. Throughout the paper, we assume that both d_θ and d_g are fixed, and the sample size n is large. A common way to estimate the unknown parameter of interest relies on the GMM estimator $\hat{\theta}_n$ solution of

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} \left(\frac{1}{n} \sum_{i=1}^{n} g(X_i, \theta) \right)' S_n \left(\frac{1}{n} \sum_{i=1}^{n} g(X_i, \theta) \right), \tag{1}$$

where S_n denotes a sequence of positive definite symmetric $d_g \times d_g$ matrices that converge in probability to a positive definite symmetric matrix S_0.

Example 1. Consider the linear regression model,

$$Y_i = W_i'\theta_0 + U_i,$$

$i = 1, \ldots, n$, where $Y_i \in \mathbb{R}$, and the regressor $W_i = (W_{i,1}, \ldots, W_{i,d_\theta})' \in \mathbb{R}^{d_\theta}$ may be correlated to the error term $U_i \in \mathbb{R}$. Suppose that there exists a d_g-dimensional random vector of instruments $Z_i = (Z_{i,1}, \ldots, Z_{i,d_g})'$, with $E[Z_{i,j}U_i] = 0$, for $j = 1, \ldots, d_g$. Then, we can easily verify that θ_0 satisfies the moment conditions $E[g(X_i, \theta_0)] = 0$, with $X_i = (Y_i, W_i', Z_i')'$ and $g(X_i, \theta) = \big((Y_i - W_i'\theta)Z_{i,1}, \ldots, (Y_i - W_i'\theta)Z_{i,d_g}\big)'$.

Under some regularity conditions, $\hat{\theta}_n$ is a consistent estimator of θ_0 with normal limit distribution; see, e.g., Hansen (1982). Let $A = \{i : \theta_{0,i} \neq 0\}$ denotes the set of the nonzero coefficients of $\theta_0 = (\theta_{0,1}, \ldots, \theta_{0,d_\theta})'$. Assume that the number of elements in set A is $|A| = d_A < d_\theta$. Similarly, let $\hat{A}_n = \{i : \hat{\theta}_{n,i} \neq 0\}$, where $\hat{\theta}_n = (\hat{\theta}_{n,1}, \ldots, \hat{\theta}_{n,d_\theta})'$. Then, in general $|\hat{A}_n| = d_\theta \neq d_A$. Thus, in spite of a consistent estimate of the unknown parameter, GMM estimators do not perform variable selection. To overcome this problem in iid linear regression models, recent research proposes penalized least squares estimators that combine consistent (and efficient) parameter estimation and variable selection in one step. Examples of penalized least squares estimators include bridge estimators with L_q penalty for $0 < q < 1$ (see, e.g., Frank and Friedman, 1993, Fu, 1998, and Knight and Fu, 2000), and adaptive lasso estimators (see, e.g., Zou, 2006). By extending these estimators to our setting, we introduce the penalized GMM estimators

$$\check{\theta}_n = \arg\min_{\theta \in \Theta} \left(\frac{1}{n} \sum_{i=1}^{n} g(X_i, \theta) \right)' S_n \left(\frac{1}{n} \sum_{i=1}^{n} g(X_i, \theta) \right) + \frac{\lambda_n}{n} \sum_{i=1}^{d_\theta} \lambda_{n,i}^{\gamma} |\theta_i|^q, \tag{2}$$

where $\lambda_{n,i} = 1/|\hat{\theta}_{n,i}|$, λ_n is a tuning parameter, and for γ and q we consider two cases: either (i) $\gamma = 0$ and $0 < q < 1$ (bridge estimators) or (ii) $\gamma = q = 1$ (adaptive lasso estimators). As pointed out in Zou (2006), in the definition of the adaptive lasso we could also consider the more general case $q = 1$ and $\gamma > 0$. For brevity, we only consider the case $\gamma = q = 1$, however similar results established in this paper can be obtained also when $q = 1$ and $\gamma > 0$. In the next section, we derive the asymptotic properties of the penalized GMM estimators.

2.2 Penalized GMM Estimators and Oracle Properties

In iid linear regression models, bridge estimators with L_q penalty for $0 < q < 1$, and adaptive lasso estimators possess the so-called oracle properties. More precisely, they simultaneously perform correct variable selection, and provide efficient estimates of the nonzero coefficients as if only the relevant variables had been included in the model. In this section, we show that the penalized GMM estimators introduced in (2) feature these properties also in our general setting. First, we introduce some notation. Let θ_0^A and $\theta_0^{A^c}$ denote the sub-vectors of the nonzero and zero coefficients of θ_0, respectively. Similarly, let $\check{\theta}_n^A$ and $\check{\theta}_n^{A^c}$ denote the penalized GMM estimators of θ_0^A and $\theta_0^{A^c}$, respectively. Finally, let $\check{A}_n = \{i : \check{\theta}_{n,i} \neq 0\}$, where $\check{\theta}_n = (\check{\theta}_{n,1}, \ldots, \check{\theta}_{n,d_\theta})'$. Before presenting the asymptotic properties of penalized GMM estimators, we introduce following assumptions.

Assumption 1.

(a) $(X_1, \ldots, X_n)$ *are iid observations of* $X \sim F$*, for some probability distribution* F*.*

(b) θ_0 *is the unique solution in* Θ *of* $E[g(X, \theta_0)] = 0$*,* Θ *is compact, and* θ_0 *is an interior point of* Θ*.*

(c) $g(X, \theta)$ *is Lipschitz continuous on* Θ*, i.e.,* $\|g(X, \theta_1) - g(X, \theta_2)\| \leq L\|\theta_1 - \theta_2\|$ *a.s. for all* $\theta_1, \theta_2 \in \Theta$*, for some constant* L*. Also* $\frac{\partial}{\partial \theta} g(X, \theta)$ *is Lipschitz continuous on* Θ*.*

(d) For some $r \geq 2$*,* $E[\|g(X, \theta)\|^{2r}] < \infty$*, and* $E[\|\frac{\partial}{\partial \theta} g(X, \theta)\|^r] < \infty$*, for all* $\theta \in \Theta$*.*

(e) (i) $\Omega_0 = E[g(X, \theta_0) g(X, \theta_0)']$ *is positive definite. (ii)* $D_0 = E[\frac{\partial}{\partial \theta} g(X, \theta)]$ *is of full rank. (iii)* S_n *converges in probability to a positive definite symmetric matrix* S_0*.*

Assumption 1 provides a set of conditions that are typically required for the consistency and asymptotic normality of GMM estimators and bootstrap approximations. The oracle properties of penalized GMM estimators are summarized in the next lemma.

Lemma 1. *Suppose that Assumption 1 holds. Furthermore, consider the covariance matrix* $V_0 = (D_0' S_0 D_0)^{-1} D_0' S_0 \Omega_0 S_0 D_0 (D_0' S_0 D_0)^{-1}$*. If either (i)* $\lambda_n / \sqrt{n^q} \to \infty$ *and* $\lambda_n / \sqrt{n} \to 0$*, for* $\gamma = 0$ *and* $0 < q < 1$*, or (ii)* $\lambda_n \to \infty$ *and* $\lambda_n / \sqrt{n} \to 0$*, for* $\gamma = q = 1$*, then, as* $n \to \infty$*,*

(I) $\sqrt{n} \check{\theta}_n^{A^c}$ *converges in probability to* 0*.*

(II) The law of $\sqrt{n} \left(\check{\theta}_n^A - \theta_0^A \right)$ *converges weakly to normal with mean* 0 *and covariance* V_0^A*, where* V_0^A *is the sub-matrix of* V_0 *for the true subset model of nonzero coefficients of* θ_0*.*

(III) $\lim_{n \to \infty} P(\check{A}_n = A) = 1$*.*

Statement (III) of Lemma 1 establishes that the penalized GMM estimators asymptotically identify the sub-vector of the nonzero coefficients of θ_0. Furthermore, statement (II) shows that the penalized GMM estimators for the nonzero coefficients has the same efficiency of the GMM estimator based on the true subset model. In particular, by considering a sequence S_n that converges in probability to Ω_0^{-1}, for instance $S_n = \left(\frac{1}{n}\sum_{i=1}^n g(X_i, \hat{\theta}_n) g(X_i, \hat{\theta}_n)'\right)^{-1}$, we can note that the penalized GMM estimators for the nonzero coefficients achieve the same efficiency of the (efficient) two-step GMM estimator for the true subset model.

Remark 1. The results in Lemma 1 are subcases of findings established in Caner (2009) and Caner and Zhang (2014). Indeed, Lemma 1 can be extended also to high-dimensional settings where d_θ and d_g may depend on the sample size n. For instance, Caner and Zhang (2014) recently establish oracle properties of adaptive elastic net estimators in moment condition models when $d_\theta \to \infty$ and $d_g \to \infty$, but $d_g/n \to 0$.

Remark 2. To prove parts (I) and (II) of Lemma 1, we extend the approach adopted in the proof of Theorem 2 in Zou (2006) to our nonlinear moment condition model. More precisely, first we show that $\sqrt{n}(\check{\theta}_n - \theta_0)$ minimizes a particular random process. Then, we compute the limit and the minimum of this random process. Finally, we apply results in Geyer (1994). On the other hand, to prove part (III) of Lemma 1, we use similar arguments adopted in the proof of Lemma 5 in Fan and Peng (2004).

3 Bootstrap for Penalized GMM Estimators

Since in our setting we do not have parametric information on the data generating process, the standard approach to bootstrapping is the nonparametric bootstrap with recentered moment conditions proposed in Hall and Horowitz (1996), Andrews (2002), and Camponovo (2016), among others. The nonparametric bootstrap constructs random samples $(X_1^*, \ldots, X_n^*)$ by selecting from $(X_1, \ldots, X_n)$ with uniform weight $1/n$ with replacement. Furthermore, we replace the moment function $g(X_i^*, \theta)$ with $g^*(X_i^*, \theta) = g(X_i^*, \theta) - \frac{1}{n}\sum_{i=1}^n g(X_i, \check{\theta}_n)$. The recentering ensures that the bootstrap moments $E^*[g(X_i^*, \check{\theta}_n)] = 0$, when $\theta = \check{\theta}_n$, which mimics the population moments $E[g(X_i, \theta)] = 0$, when $\theta = \theta_0$. Consider the bootstrap penalized GMM estimators

$$\check{\theta}_n^* = \arg\min_{\theta \in \Theta} \left(\frac{1}{n}\sum_{i=1}^n g^*(X_i^*, \theta)\right)' S_n \left(\frac{1}{n}\sum_{i=1}^n g^*(X_i^*, \theta)\right) + \frac{\lambda_n}{n}\sum_{i=1}^{d_\theta} (\lambda_{n,i}^*)^\gamma |\theta_i|^q, \tag{3}$$

where $\lambda_{n,i}^* = 1/|\hat{\theta}_{n,i}^*|$, $\hat{\theta}_n^* = (\hat{\theta}_{n,1}^*, \ldots, \hat{\theta}_{n,d_\theta}^*)'$ is the GMM estimator solution of (1) based on the bootstrap sample $(X_1^*, \ldots, X_n^*)$, λ_n is a tuning parameter, and either (i) $\gamma = 0$ and $0 < q < 1$, or (ii) $\gamma = q = 1$. The nonparametric bootstrap approximates the sampling distribution of $\sqrt{n}(\check{\theta}_n - \theta_0)$ with the conditional distribution of $\sqrt{n}(\check{\theta}_n^* - \check{\theta}_n)$ given the observations $(X_1, \ldots, X_n)$.

Let $\check{\theta}_n^{A*}$ and $\check{\theta}_n^{A^c*}$ denote the bootstrap penalized GMM estimators of θ_0^A and $\theta_0^{A^c}$, respectively. In the next theorem, we prove the validity of the bootstrap approximation.

Theorem 1. *Suppose that Assumption 1 holds. If either (i) $\lambda_n/\sqrt{n^q} \to \infty$ and $\lambda_n/\sqrt{n} \to 0$, for $\gamma = 0$ and $0 < q < 1$, or (ii) $\lambda_n \to \infty$ and $\lambda_n/\sqrt{n} \to 0$, for $\gamma = q = 1$, then, as $n \to \infty$,*

(I) $\sqrt{n}\check{\theta}_n^{A^c*}$ converges in conditional probability to 0.

(II) The conditional law of $\sqrt{n}\left(\check{\theta}_n^{A*} - \check{\theta}_n^{A}\right)$ converges weakly to normal with mean 0 and covariance V_0^A.

The results in Theorem 1 show that the nonparametric bootstrap provides a valid approach for approximating the sampling distribution of penalized GMM estimators. Furthermore, by adapting the results in Corollary 3.2 in Chatterje and Lahiri (2011) to our setting, we can show that the nonparametric bootstrap can also be applied for the construction of confidence sets for the unknown parameter of interest.

Remark 3. To prove Theorem 1, we extend the approach adopted in the proof of Theorem 2 in Zou (2006) to our bootstrap nonlinear moment condition model. More precisely, first we show that $\sqrt{n}(\check{\theta}_n^* - \check{\theta}_n)$ minimizes a particular random process. Then, we compute the limit of this process. To this end, we consider the conditional probability given the sample $(X_1, \ldots, X_n)$, and compute the limit by successively conditioning on a sequence of samples, as $n \to \infty$. Finally, we compute the minimum of this random process and apply results in Geyer (1994).

4 Selection of the Tuning Parameter

The accuracy of penalized GMM estimators may heavily depend on the selection of λ_n. By adapting the nonparametric bootstrap, we can introduce a data-driven method for the selection of this tuning parameter in the spirit of Hall et al. (2009), Chatterje and Lahiri (2011) and Camponovo (2015). The key idea of our approach is to select the optimal tuning parameter that minimizes the estimated mean squared error of the adaptive lasso estimator $\check{\theta}_n$.

To this end, also in this case we replace the moment function $g(X_i^*, \theta)$ with $\tilde{g}^*(X_i^*, \theta) = g(X_i^*, \theta) - \frac{1}{n}\sum_{i=1}^n g(X_i, \hat{\theta}_n)$. Note that instead of recentering with respect to the penalized GMM estimator $\check{\theta}_n$, we recenter with respect to the GMM estimator $\hat{\theta}_n$ that does not depend on the selection of λ_n. Furthermore, we introduce the recentered bootstrap penalized GMM estimators

$$\tilde{\theta}_n^*(\lambda_n) = \arg\min_{\theta\in\Theta} \left(\frac{1}{n}\sum_{i=1}^n \tilde{g}^*(X_i^*, \theta)\right)' S_n \left(\frac{1}{n}\sum_{i=1}^n \tilde{g}^*(X_i^*, \theta)\right) + \frac{\lambda_n}{n}\sum_{i=1}^{d_\theta} (\lambda_{n,i}^*)^\gamma |\theta_i - \hat{\theta}_{n,i} I(\check{\theta}_{n,i} = 0)|^q, \quad (4)$$

where $\lambda^*_{n,i} = 1/|\hat{\theta}^*_{n,i}|$, λ_n is a tuning parameter, I is the indicator function and either (i) $\gamma = 0$ and $0 < q < 1$, or (ii) $\gamma = q = 1$.

In the penalization term in (4) we recenter with respect to $\hat{\theta}_{n,i} I(\check{\theta}_{n,i} = 0)$. This recentering has no impact on the nonzero coefficients of θ_0 for large n. Indeed, if $\theta_{0,i} \neq 0$, then $|\check{\theta}_{n,i}| > 0$ for large n with high probability, and consequently $\theta_i - \hat{\theta}_{n,i} I(\check{\theta}_{n,i} = 0) = \theta_i$. On the other hand, if $\theta_{0,i} = 0$, then $\check{\theta}_{n,i} = 0$ for large n with high probability, and consequently $\theta_i - \hat{\theta}_{n,i} I(\check{\theta}_{n,i} = 0) = \theta_i - \hat{\theta}_{n,i}$. Therefore, the penalization term in (4) shrinks the bootstrap estimates of zero coefficients of θ_0 to the GMM estimates. This adjustment exactly mimics the standard penalization term that shrinks the estimates of zero coefficients of θ_0 to 0; see, Camponovo (2015) for further details on this approach. Finally, we estimate the mean squared error $E[\|\check{\theta}_n - \theta_0\|^2]$ by

$$\phi(\lambda_n) = E^*[\|\tilde{\theta}^*_n(\lambda_n) - \hat{\theta}_n\|^2], \tag{5}$$

where E^* denotes the expectation with respect to the distribution of the bootstrap sample conditional on the original sample, and select the optimal value $\hat{\lambda}_n$ that minimizes (5). In the Monte Carlo analysis presented in Section 5, we study through Monte Carlo simulations the accuracy of the nonparametric bootstrap approach combined with this data-driven method for the selection of the tuning parameter.

5 Monte Carlo

In this section, we study the accuracy of inference based on GMM estimators with normal approximation, and adaptive lasso penalized GMM estimators with the nonparametric bootstrap. For the adaptive lasso penalized GMM estimators, we select the tuning parameters λ_n according to the data-driven method introduced in Remark 4. The number of random samples is $N = 2000$, and the number of bootstrap replications is $B = 299$.

In the first exercise, we consider the setting introduced in Example 1 with $n = 150, 300$, $d_\theta = 20$, and $d_g = 30, 40$. The true value θ_0 contains five large coefficients $\theta_{0,1} = \cdots = \theta_{0,5} = 1$, five moderate coefficients $\theta_{0,6} = \cdots = \theta_{0,10} = 0.5$, and ten noise coefficients $\theta_{0,11} = \cdots = \theta_{0,20} = 0$. For the covariates and the error terms, we assume $W_{i,t} \sim N(0,1)$, and $U_i \sim N(0, \sigma_j)$, $j = 1, 2$, with $\sigma_1 = 1$ and $\sigma_2 = \frac{1}{d_\theta} \sum_{i=1}^{d_\theta} W^2_{i,t}$. Furthermore, we consider strong instruments $Z_{i,t} \sim N(0,1)$ with $E[W_{i,t} Z_{i,t}]$ uniformly selected in the interval $[0.6, 0.9]$. In Tables 1 and 2, we report the empirical coverages and the mean of the length of symmetric 0.95-confidence intervals for large, moderate and noise coefficients. For nonzero coefficients, the nonparametric bootstrap provides empirical coverages very close to 0.95, and always outperforms inference based on GMM estimator and normal approximation. For zero coefficients, the empirical coverages of normal approximation with GMM estimator are slightly smaller than 0.95. On the other hand, the nonparametric bootstrap with adaptive lasso penalized GMM estimator provide shorter confidence intervals with coverage converging to 1. Indeed, in Lemma 1 we show that the penalized GMM estimates of zero coefficients collapse to 0 asymptotically. Therefore, in this case the coverage of confidence intervals should converge to 1 as n increases; see e.g., Minnier, et al. (2011) for similar empirical findings.

In the second exercise, we consider the setting introduced in Example 1 with $n = 100$, $d_\theta = 10$, and $d_g = 20, 30$. For the covariates and the error terms we assume $W_{i,t} \sim N(0,1)$ and $U_i \sim N(0,1)$. Furthermore, also in this case, we consider strong instruments $Z_{i,t} \sim N(0,1)$ with $E[W_{i,t}Z_{i,t}]$ uniformly selected in the interval $[0.6, 0.9]$. The true value θ_0 contains three large coefficients $\theta_{0,1} = \theta_{0,2} = \theta_{0,3} = 1$, three moderate coefficients $\theta_{0,4} = \theta_{0,5} = \theta_{0,6} = 0.5$, three noise coefficients $\theta_{0,7} = \theta_{0,8} = \theta_{0,9} = 0$, and $\theta_{0,10} = c$, with $c \in [0, 4/\sqrt{n}]$. In Tables 3 and 4, we report the empirical rejection frequencies of the null hypothesis $H_0 : \beta_{0,10} = 0$, versus the alternative $H_1 : \beta_{0,10} \neq 0$, for $c \in [0, 4/\sqrt{n}]$, and significance level 0.05. When $c = 0$, the rejection frequencies using normal approximation with GMM estimator are slightly larger than the significance level. On the other hand, in line with the previous exercise, the nonparametric bootstrap with adaptive lasso penalized GMM estimator provide rejection frequencies that tend to be quite close to 0. As expected, when $c > 0$ the power increases. The normal approximation with GMM estimator imply larger rejection frequencies. However, the difference in power with the nonparametric bootstrap is always smaller than 0.10.

6 Conclusions

The GMM provides a powerful tool for introducing statistical inference in several economic and financial models that are specified by some moment conditions. However, in spite of a consistent (and efficient) estimate of the parameter of interest, GMM estimators do not perform variable selection. To overcome this problem, recent research proposes penalized GMM estimators that may achieve the so-called oracle properties. In this paper, we study the validity of bootstrap methods in approximating the sampling distribution of penalized GMM estimators with oracle properties. More precisely, we focus on bridge estimators with L_q penalty for $0 < q < 1$, and adaptive lasso estimators. In particular, we show that the nonparametric bootstrap with recentered moment conditions provides a valid method for approximating the sampling distribution of penalized GMM estimators. Furthermore, using the bootstrap approach we also propose a data-driven method for the selection of tuning parameters in the penalization terms. Monte Carlo simulations confirm the reliability and accuracy of the bootstrap procedure.

References

[1] Andrews, D.W.K. (2002). Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators. *Econometrica*, **70**, 119-162.

[2] Belloni, A. Chernozhukov, V. and Hansen, C. (2014a). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, **81**, 608-650.

[3] Belloni, A. Chernozhukov, V. and Hansen, C. (2014b). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, **28**, 29-50.

[4] Camponovo, L., (2015). On the validity of the pairs bootstrap for lasso estimators. *Biometrika*, **102**, 981-987.

[5] Camponovo, L., (2016). Asymptotic refinements of nonparametric bootstrap for quasi-likelihood ratio tests for classes of extremum estimators. *Econometrics Journal*, **19**, 33-54. *Biometrika*, **102**, 981-987.

[6] Caner, M., (2019). Lasso-type GMM estimtor. *Econometric Theory*, **25**, 270-290.

[7] Caner, M., and Zang, H.H., (2014). Adaptive elastic net for generalized methods of moment. *Journal of Business and Economic Statistic*, **32**, 30-47.

[8] Chatterjee, A. and Lahiri, S.N., (2010). Asymptotic properties of the residual bootstrap for Lasso estimators. *Proc. Amer. Math. Soc.*, **138**, 4497-4509.

[9] Chatterjee, A., and Lahiri, S.N., (2011). Bootstrapping lasso estimators. *Journal of the American Statistical Association*, **106**, 608-625.

[10] Chatterjee, A., and Lahiri, S.N., (2013). Rates of convergence of the adaptive LASSO estimators to the oracle distribution and higher order refinements by the bootstrap. *Annals of Statistics*, **41**, 1232-1259.

[11] 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.

[12] Fan, J., and Peng, H., (2004). On nonconcave penalized likelihood with diverging number of parameters. *Annals of Statistics*, **32**, 928-961.

[13] Frank, I.E., and Friedman, J.H., (1993). A statistical view of some chemometrics regression tools (with discussion). *Technometrics*, **35**, 109-148.

[14] Fu, W.J., (1998). Penalized regressions: the bridge versus the lasso. *Journal of Computational and Graphical Statistics*, **7**, 397-416.

[15] Geyer, C., (1994). On the asymptotics of constrained M-estimation. *Annals of Statistics*, **22**, 1993-2010.

[16] Hall, A.R., (2005). *Generalized Method of Moments*. Oxford University Press, Oxford.

[17] Hall, P., and Horowitz, J., (1996). Bootstrap critical values for tests based on generalized-method-0f-moments estimators. *Econometrica*, **64**, 891-916.

[18] Hall, P., Lee, E.R. and Park, B.U., (2009). Bootstrap-based penalty choice for the lasso achieving oracle performance. *Statistica Sinica*, **19**, 449-471.

[19] Hansen, L.P., (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, **50**, 1029-1054.

[20] Knight, K., and Fu, W., (2000). Asymptotics for lasso-type estimators. *Annals of Statistics*, **28**, 1356-1378.

[21] Lockhart, T. Taylor, J., Tibshirani R.J., and Tibshirani R., (2014). A significance test for the lasso. *Annals of Statistics*, **42**, 413-468.

[22] Minnier, J., Tian, L. and Cai, T., (2011). A perturbation method for inference on regularized regression estimates. *Journal of the American Statistical Association*, **106**, 1371-1382.

[23] Taylor, J., Lockhart, R., Tibshirani R.J., and Tibshirani R., (2014). Post-selection adaptive inference for angle regression and lasso. Working paper.

[24] Tibshirani, R., (1996). Regression analysis and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, **58**, 267-288.

[25] van de Geer, S., Bühlmann, P., Ritov, Y., and Dezeure, R., (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Annals of Statistics*, **42**, 1166-1202.

[26] Zhang, C.H., and Zhang, S.S., (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society, Series B*, **76**, 217-242.

[27] Zou, H., (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, **101**, 1418-1429.

Appendix: Assumptions and Mathematical Proofs

Proof of Lemma 1: To prove Lemma 1, we extend the approach adopted in the proof of Theorem 2 in Zou (2006) to our nonlinear moment condition model. Let $J_n(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(X_i, \theta)$, and consider the random process,

$$R_n(u) = J_n(\theta_0 + u/\sqrt{n})' S_n J_n(\theta_0 + u/\sqrt{n}) - J_n(\theta_0)' S_n J_n(\theta_0) + \lambda_n \sum_{i=1}^{d_\theta} \lambda_{n,i}^{\gamma} \left[|\theta_{0,i} + u_i/\sqrt{n}|^q - |\theta_{0,i}|^q \right].$$

Note that $R_n(u)$ is minimized at $\sqrt{n}(\check{\theta}_n - \theta_0)$. By considering a Taylor expansion of $J_n(\theta_0 + u/\sqrt{n})$ around θ_0 we have,

$$J_n(\theta_0 + u/\sqrt{n}) = J_n(\theta_0) + D_n(\theta_0)u + o_p(1), \tag{6}$$

where $D_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial\theta} g(X_i, \theta)$. It turns out that using (6) we can rewrite $R_n(u)$ as

$$\begin{aligned} R_n(u) &= 2u' D_n(\theta_0)' S_n J_n(\theta_0) + u' D_n(\theta_0)' S_n D_n(\theta_0) u \\ &\quad + \lambda_n \sum_{i=1}^{d_\theta} \lambda_{n,i}^{\gamma} \left[|\theta_{0,i} + u_i/\sqrt{n}|^q - |\theta_{0,i}|^q \right] + o_p(1). \end{aligned}$$

Note that under Assumption 1, $D_n(\theta_0)$ converges in probability to D_0, while the law of $J_n(\theta_0)$ converges weakly to normal with mean 0 and variance Ω_0. Furthermore, S_n converges in probability to S_0.

Next, consider the term $\lambda_n \sum_{i=1}^{d_\theta} \lambda_{n,i}^{\gamma} [|\theta_{0,i} + u_i/\sqrt{n}|^q - |\theta_{0,i}|^q]$, when $\gamma = 0$ and $0 < q < 1$. If $\theta_{0,i} \neq 0$, then $\lambda_n [|\theta_{0,i} + u_i/\sqrt{n}|^q - |\theta_{0,i}|^q]$ converges in probability to 0, since $\lambda_n/\sqrt{n} \to 0$, as $n \to \infty$. On the other hand, if $\theta_{0,i} = 0$, then $|\theta_{0,i} + u_i/\sqrt{n}|^q - |\theta_{0,i}|^q = |u_i|^q/\sqrt{n^q}$. Let $\tilde{u} = (\tilde{u}_1, \ldots, \tilde{u}_{d_\theta})'$, with $\tilde{u}_i = u_i$ for $i \in A$, and $\tilde{u}_i = 0$, for $i \notin A$. Then, since $\lambda_n/\sqrt{n^q} \to \infty$, the limit $R(u)$ of $R_n(u)$ is given by

$$R(u) = \begin{cases} 2\tilde{u}' D_0' S_0 \omega_0 + \tilde{u}' D_0' S_0 D_0 \tilde{u}, & \text{if } u_i = 0, \text{ for } i \notin A, \\ \infty, & \text{otherwise,} \end{cases}$$

where $\omega_0 \sim N(0, \Omega_0)$. Note that the unique minimum of $R(u)$ is $((-(D_0^{A'} S_0 D_0^A)^{-1} D_0^{A'} S_0 w_0)', 0')'$, where D_0^A is the sub-matrix of D_0 for the nonzero coefficients. Therefore, using the results in Geyer (1994), parts (I) and (II) of Lemma 1 for bridge estimators with L_q penalty for $0 < q < 1$ are established.

Next, consider the term $\lambda_n \sum_{i=1}^{d_\theta} \lambda_{n,i}^{\gamma} [|\theta_{0,i} + u_i/\sqrt{n}|^q - |\theta_{0,i}|^q]$, when $\gamma = q = 1$. If $\theta_{0,i} \neq 0$, then $\lambda_n \lambda_{n,i} [|\theta_{0,i} + u_i/\sqrt{n}| - |\theta_{0,i}|]$ converges in probability to 0, since $\lambda_n/\sqrt{n} \to 0$, as $n \to \infty$. On the other hand, if $\theta_{0,i} = 0$, then $|\theta_{0,i} + u_i/\sqrt{n}| - |\theta_{0,i}| = |u_i|/\sqrt{n}$, and $\lambda_{n,i} = O_p(\sqrt{n})$. Then, since $\lambda_n \to \infty$, also in this case the limit $R(u)$ of $R_n(u)$ is given by

$$R(u) = \begin{cases} 2\tilde{u}' D_0' S_0 \omega_0 + \tilde{u}' D_0' S_0 D_0 \tilde{u}, & \text{if } u_i = 0, \text{ for } i \notin A, \\ \infty, & \text{otherwise,} \end{cases}$$

where $\omega_0 \sim N(0, \Omega_0)$. Therefore, parts (I) and (II) of Lemma 1 for adaptive lasso estimators are also established.

Next, we prove part (III) using similar arguments adopted in the proof of Lemma 5 in Fan and Peng (2004). Let

$$Q(\theta) = \left(\frac{1}{n}\sum_{t=1}^{n} g(X_t, \theta)\right)' S_n \left(\frac{1}{n}\sum_{t=1}^{n} g(X_t, \theta)\right) + \frac{\lambda_n}{n}\sum_{t=1}^{d_\theta} \lambda_{n,t}^{\gamma} |\theta_t|^q.$$

With some abuse of notation, we write $\theta = (\theta^{A'}, \theta^{A^{c'}})'$. We show that with probability tending to 1, for any $\tilde{\theta}_n^A$ satisfying $\|\tilde{\theta}_n^A - \theta_0^A\| = O_p(1/\sqrt{n})$ and any constant $c > 0$,

$$Q((\tilde{\theta}_n^{A'}, 0')') = \min_{\|\theta^{A^c}\| \le c/\sqrt{n}} Q((\tilde{\theta}_n^{A'}, \theta^{A^{c'}})').$$

To this end, for $i \notin A$ consider

$$\begin{aligned} \frac{\partial Q(\theta)}{\partial \theta_i} &= 2\left(\frac{1}{n}\sum_{t=1}^{n} \frac{\partial}{\partial \theta_i} g(X_t, \theta)\right)' S_n \left(\frac{1}{n}\sum_{t=1}^{n} g(X_t, \theta)\right) + \frac{\lambda_n}{n} \lambda_{n,i}^{\gamma} \operatorname{sgn}(\theta_i) q |\theta_i|^{q-1}, \\ &= I_1 + I_2. \end{aligned}$$

Note that for $\|\tilde{\theta}_n^A - \theta_0^A\| = O_p(1/\sqrt{n})$, and $\|\theta^{A^c}\| \le c/\sqrt{n}$, then $I_1 = O_p(1/\sqrt{n})$.

Next, consider the second term I_2, when $\gamma = 0$ and $0 < q < 1$. Note that $|\theta_i| \le c/\sqrt{n}$. Therefore, since $\lambda_n/\sqrt{n^q} \to +\infty$, the dominant term is I_2. Thus, the sign of θ_i determines the sign of $\partial Q(\theta)/\partial \theta_i$. More precisely, we have $\partial Q(\theta)/\partial \theta_i < 0$, when $-c/\sqrt{n} < \theta_i < 0$, and $\partial Q(\theta)/\partial \theta_i > 0$, when $0 < \theta_i < c/\sqrt{n}$. This concludes the proof of part (III) for bridge estimators with L_q penalty for $0 < q < 1$.

Next, consider the second term I_2, when $\gamma = q = 1$. Since $\lambda_n \to +\infty$, it turns out that the dominant term is I_2. Thus, the sign of θ_i determines the sign of $\partial Q(\theta)/\partial \theta_i$. More precisely, also in this case we have $\partial Q(\theta)/\partial \theta_i < 0$, when $-c/\sqrt{n} < \theta_i < 0$, and $\partial Q(\theta)/\partial \theta_i > 0$, when $0 < \theta_i < c/\sqrt{n}$. This concludes the proof of part (III) for adaptive lasso estimators.

Proof of Theorem 1: To prove Theorem 1, we extend the approach adopted in the proof of Theorem 2 in Zou (2006) to our bootstrap nonlinear moment condition model. More precisely, let $J_n^*(\theta) = \left(\frac{1}{\sqrt{n}}\sum_{t=1}^n g^*(X_t^*, \theta)\right)$, and consider the process,

$$R_n^*(u) = J_n^*(\check{\theta}_n + u/\sqrt{n})' S_n J_n^*(\check{\theta}_n + u/\sqrt{n}) - J_n^*(\check{\theta}_n)' S_n J_n^*(\check{\theta}_n) + \lambda_n \sum_{i=1}^{d_\theta} (\lambda_{n,i}^*)^\gamma \left[|\check{\theta}_{n,i} + u_i/\sqrt{n}|^q - |\check{\theta}_{n,i}|^q\right].$$

Note that $R_n^*(u)$ is minimized at $\sqrt{n}(\tilde{\theta}_n^* - \check{\theta}_n)$. By considering a Taylor expansion of $J_n^*(\check{\theta}_n + u/\sqrt{n})$ around $\check{\theta}_n$, and by Assumption 1, we can rewrite $R_n^*(u)$ as

$$\begin{aligned} R_n^*(u) &= 2u' D_n^*(\check{\theta}_n)' S_n J_n^*(\check{\theta}_n) + u' D_n^*(\check{\theta}_n)' S_n D_n^*(\check{\theta}_n) u \\ &\quad + \lambda_n \sum_{i=1}^{d_\theta} (\lambda_{n,i}^*)^\gamma \left[|\check{\theta}_{n,i} + u_i/\sqrt{n}|^q - |\check{\theta}_{n,i}|^q\right] + o_p(1), \end{aligned}$$

where $D_n^*(\theta) = \frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial\theta} g(X_i^*, \theta)$. Note that under Assumption 1, $D_n^*(\check{\theta}_n)$ converges in conditional probability to D_0. Furthermore, the conditional law of $J_n^*(\check{\theta}_n)$ converges weakly to normal with mean 0 and variance Ω_0.

Next, consider the term $\lambda_n \sum_{i=1}^{d_\theta} (\lambda_{n,i}^*)^\gamma \left[|\check{\theta}_{n,i} + u_i/\sqrt{n}|^q - |\check{\theta}_{n,i}|^q\right]$, when $\gamma = 0$ and $0 < q < 1$. If $\theta_{0,i} \neq 0$, then $\check{\theta}_{n,i} \neq 0$ for n large. Therefore, $\lambda_n[|\check{\theta}_{n,i} + u_i/\sqrt{n}|^q - |\check{\theta}_{n,i}|^q]$ converges in probability to 0, since $\lambda_n/\sqrt{n} \to 0$, as $n \to \infty$. On the other hand, if $\theta_{0,i} = 0$, then $\check{\theta}_{n,i} = 0$ for n large. Let $\tilde{u} = (\tilde{u}_1, \ldots, \tilde{u}_{d_\theta})'$, with $\tilde{u}_i = u_i$ for $i \in A$, and $\tilde{u}_i = 0$, for $i \notin A$. Then, since $\lambda_n/\sqrt{n^q} \to \infty$, the limit $R(u)$ of $R_n^*(u)$ is given by

$$R(u) = \begin{cases} 2\tilde{u}' D_0' S_0 \omega_0 + \tilde{u}' D_0' S_0 D_0 \tilde{u}, & \text{if } u_i = 0, \text{ for } i \notin A, \\ \infty, & \text{otherwise,} \end{cases}$$

where $\omega_0 \sim N(0, \Omega_0)$.

Finally, consider the term $\lambda_n \sum_{i=1}^{d_\theta} (\lambda_{n,i}^*)^\gamma \left[|\check{\theta}_{n,i} + u_i/\sqrt{n}|^q - |\check{\theta}_{n,i}|^q\right]$, when $\gamma = q = 1$. If $\theta_{0,i} \neq 0$, then $\check{\theta}_{n,i} \neq 0$ for n large. Therefore, since $\lambda_n/\sqrt{n} \to 0$, it turns out that $\lambda_n \lambda_{n,i}^* \{|\check{\theta}_{n,i} + u_i/\sqrt{n}| - |\check{\theta}_{n,i}|\}$ converges in conditional probability to 0. On the other hand, if $\theta_{0,i} = 0$, then $\check{\theta}_{n,i} = 0$ for n large. Then, since $\lambda_n \to \infty$, also in this case the limit $R(u)$ of $R_n^*(u)$ is given by

$$R(u) = \begin{cases} 2\tilde{u}' D_0' S_0 \omega_0 + \tilde{u}' D_0' S_0 D_0 \tilde{u}, & \text{if } u_i = 0, \text{ for } i \notin A, \\ \infty, & \text{otherwise,} \end{cases}$$

where $\omega_0 \sim N(0, \Omega_0)$. Note that the unique minimum of $R(u)$ is $((-(D_0^{A'} S_0 D_0^A)^{-1} D_0^{A'} S_0 w_0)', 0')'$. Therefore, using the results in Geyer (1994), Theorem 1 is established.

$n = 150$	$\theta_{0,i} = 1$		$\theta_{0,i} = 0.5$		$\theta_{0,i} = 0$	
	σ_1	σ_2	σ_1	σ_2	σ_1	σ_2
Norm GMM	0.938	0.933	0.939	0.936	0.937	0.935
	(0.403)	(0.407)	(0.404)	(0.408)	(0.404)	(0.408)
Boot AL	0.957	0.956	0.950	0.948	0.982	0.980
	(0.446)	(0.454)	(0.445)	(0.453)	(0.153)	(0.161)
$n = 300$	$\theta_i = 1$		$\theta_i = 0.5$		$\theta_i = 0$	
	σ_1	σ_2	σ_1	σ_2	σ_1	σ_2
Norm GMM	0.940	0.939	0.939	0.937	0.942	0.940
	(0.279)	(0.285)	(0.278)	(0.284)	(0.278)	(0.284)
Boot AL	0.956	0.954	0.948	0.947	0.989	0.987
	(0.299)	(0.307)	(0.299)	(0.307)	(0.088)	(0.093)

Table 1: **Empirical Coverages and Confidence Interval Lengths.** Empirical coverages and in brackets confidence interval lengths, using GMM estimators and normal approximation (denoted by Norm GMM), and adaptive lasso penalized GMM estimators and the nonparametric bootstrap (denoted by Boot AL). The sample size is $n = 150, 300$. The number of moment conditions is $d_g = 30$.

$n = 150$	$\theta_{0,i} = 1$		$\theta_{0,i} = 0.5$		$\theta_{0,i} = 0$	
	σ_1	σ_2	σ_1	σ_2	σ_1	σ_2
Norm GMM	0.937	0.933	0.938	0.935	0.936	0.933
	(0.383)	(0.387)	(0.384)	(0.388)	(0.383)	(0.388)
Boot AL	0.955	0.951	0.945	0.945	0.983	0.981
	(0.418)	(0.427)	(0.418)	(0.427)	(0.141)	(0.149)
$n = 300$	$\theta_i = 1$		$\theta_i = 0.5$		$\theta_i = 0$	
	σ_1	σ_2	σ_1	σ_2	σ_1	σ_2
Norm GMM	0.941	0.939	0.940	0.939	0.943	0.942
	(0.267)	(0.273)	(0.266)	(0.272)	(0.266)	(0.273)
Boot AL	0.952	0.951	0.946	0.944	0.990	0.988
	(0.284)	(0.293)	(0.284)	(0.292)	(0.082)	(0.086)

Table 2: **Empirical Coverages and Confidence Interval Lengths.** Empirical coverages and in brackets confidence interval lengths, using GMM estimators and normal approximation (denoted by Norm GMM), and adaptive lasso penalized GMM estimators and the nonparametric bootstrap (denoted by Boot AL). The sample size is $n = 150, 300$. The number of moment conditions is $d_g = 40$.

$n = 100$	$\theta_{0,10} = 0$	$\theta_{0,10} = 1/\sqrt{n}$	$\theta_{0,10} = 2/\sqrt{n}$	$\theta_{0,10} = 3/\sqrt{n}$	$\theta_{0,10} = 4/\sqrt{n}$
Norm GMM	0.061	0.149	0.402	0.721	0.912
Boot AL	0.019	0.105	0.337	0.629	0.849

Table 3: **Empirical Rejection Frequencies.** Empirical rejection frequencies of $H_0 : \theta_{0,10} = 0$, using GMM estimators and normal approximation (denoted by Norm GMM), and adaptive lasso penalized GMM estimators and the nonparametric bootstrap (denoted by Boot AL). The sample size is $n = 100$, and the number of moment conditions is $d_g = 20$.

$n = 100$	$\theta_{0,10} = 0$	$\theta_{0,10} = 1/\sqrt{n}$	$\theta_{0,10} = 2/\sqrt{n}$	$\theta_{0,10} = 3/\sqrt{n}$	$\theta_{0,10} = 4/\sqrt{n}$
Norm GMM	0.063	0.162	0.443	0.752	0.940
Boot AL	0.018	0.116	0.351	0.659	0.883

Table 4: **Empirical Rejection Frequencies.** Empirical rejection frequencies of $H_0 : \theta_{0,10} = 0$, using GMM estimators and normal approximation (denoted by Norm GMM), and adaptive lasso penalized GMM estimators and the nonparametric bootstrap (denoted by Boot AL). The sample size is $n = 100$, and the number of moment conditions is $d_g = 30$.