On the Asymptotic Size of Leamer's Extreme Bounds Analysis

Dimitris Gkountanis[a] Byunghoon Kang[b]

August 12, 2026

Abstract

We revisit Leamer's Extreme Bounds Analysis (EBA), a method for assessing whether a coefficient of interest is robust across a range of model specifications. We analyze the EBA in a classical hypothesis testing framework, and show that it is related to a test based on the minimum absolute t-statistic across the submodels. We show that the original EBA, which labels a variable "robust" only if its estimated coefficient is statistically significant with the same sign in all models considered, is asymptotically conservative when the standard normal critical value (e.g., 1.96) is used. We propose a modified EBA procedure that utilizes a restricted wild bootstrap to obtain critical values from the sampling distribution of the minimum absolute t-statistic, and show that it provides asymptotically valid size control. Monte Carlo experiments confirm the theoretical findings, and we revisit Levine and Renelt (1992, AER) as an illustrative example.

JEL classification: C12, C15, C21, C52

Keywords: extreme bounds analysis, growth regressions, restricted wild bootstrap, sensitivity analysis, specification search

1 Introduction

Sensitivity analysis is a core practice in applied economics used to evaluate how changes in the assumptions or variables of a model affect the results, and it helps researchers assess the robustness of their models. Extreme Bounds Analysis (EBA), initially developed by Leamer (1983, 1985), is a specific form of sensitivity analysis designed to address issues of specification search in econometrics.[1] In regression analysis, when there are many potential explanatory variables, the central idea of EBA is to test whether the sign of the coefficient associated with a variable of interest is robust

[a]Department of Economics, Lancaster University, Lancaster, LA1 4YX, UK. Email: jimright1789@yahoo.gr

[b]Department of Economics, Lancaster University, Lancaster, LA1 4YX, UK. Email: b.kang1@lancaster.ac.uk. Corresponding Author. All errors are our own.

[1]EBA has been used in many different fields of economics and social sciences. For example, economists have used it to examine the determinants of long-term economic growth across countries (Levine and Renelt (1992), Sala-i-Martin (1997), Sturm and de Haan (2005)). Additional applications include studying the relationship between wage inequality and crime rates (Fowles and Merva (1996)), the effects of concealed weapons laws (Bartley and Cohen (1998)), lending decisions by the International Monetary Fund (Moser and Sturm (2011)), and the environmental Kuznets curve (Yang et al. (2015)). Available statistical packages for conducting EBA include the `eba` (Impavido (1998)) for Stata, and `ExtremeBounds` (Hlavac (2016)) for R.

across various model specifications, i.e., whether the sign of the estimated coefficient remains the same regardless of which control variables are included.

Not surprisingly, as discussed in the literature, the EBA criterion can be too restrictive (Levine and Renelt 1992, Levine and Zervos 1993, and Sala-i-Martin 1997) as it considers a variable "robust" only if the estimated coefficient is statistically significant for *all* models considered. Many papers in the literature have generally concluded that most, if not all, examined variables are not "robust". The EBA procedures determine the variable to be "fragile" if a single regression finds a variable insignificant, even if all other specifications find a variable significant. This problem is well recognized, and solutions have been proposed to restrict attention to better fitting models, such as the "reasonable extreme bounds" proposed by Granger and Uhlig (1990).

In our paper, we revisit the EBA analysis and establish its connection with the classical hypothesis testing framework. The main contribution of this paper is the novel observation that the EBA procedure is related to considering the minimum absolute t-statistic across submodels together with an appropriate critical value. This observation is useful because it turns an informal sensitivity analysis rule into a well-defined test statistic, so that the size of the EBA can be studied within the standard framework. We show that the asymptotic size of the original EBA procedure with the standard normal critical value (e.g., 1.96 for the 5% significance level) is conservative, i.e., it is a testing procedure that has an asymptotic size substantially below the desired significance level.

Within this framework, the early critique raised by Sala-i-Martin (1997) is related to the use of "minimum" t-statistics that have low power. In this paper, we emphasize that an additional source of conservativeness also comes from the use of the standard normal critical value that is calculated for a single t-test.

We propose the modified EBA procedure that uses critical values from the distribution of the minimum absolute t-statistic, which are typically lower than the conventional 1.96 value (ranging from 1.16 to 1.69 in our simulation settings). Because the limiting distribution of the joint t-statistics depends on the unknown dependence across submodels, we obtain the critical values by using a bootstrap. In particular, the critical values are implemented using a restricted wild bootstrap that imposes the null hypothesis while preserving the omitted variable bias structure, following the well-established principle that bootstrap resampling should reflect the null hypothesis being tested (Hall and Wilson (1991), Davidson and MacKinnon (1999)).[2] Because the EBA criterion is based on the extremum functional of a vector of studentized statistics, our analysis is also related to inference based on maxima of statistics in the data-snooping and multiple-testing literature (White (2000), Romano and Wolf (2005)). To the best of our knowledge, analyzing the EBA procedure based on the minimum of absolute t-statistics formally within a frequentist testing setup, and examining its asymptotic size properties, are new to the literature.

[2] The wild bootstrap (Wu (1986), Liu (1988), Mammen (1993)) is a natural choice in our setting because it accommodates heteroskedasticity of unknown form and it has become a standard tool for bootstrap inference in regression models (MacKinnon (2006), Davidson and Flachaire (2008), Davidson and MacKinnon (2010)). In settings with clustered dependent data, see also Cameron, Gelbach and Miller (2008), Djogbenou, MacKinnon and Nielsen (2019), MacKinnon, Nielsen and Webb (2023), although we focus on the i.i.d. case. See Horowitz (2001) and MacKinnon (2009) for general treatments of the bootstrap and bootstrap testing.

By using smaller critical values from the minimum of the t-statistics, the modified EBA yields narrower extreme bounds, which in turn decreases the chance of including 0, thereby increasing the likelihood that a variable is classified as "robust" while maintaining appropriate statistical size. Consequently, some variables previously categorized as "fragile" in Levine and Renelt (1992) are classified as "robust" with the modified EBA procedures in our empirical studies.

We recognize that EBA and its variants are not the only approaches available for choosing the most relevant predictors or incorporating model uncertainty; model selection and specification search have a long tradition in econometrics. Alternative approaches include various frequentist and Bayesian model averaging techniques (Hjort and Claeskens (2003), Sala-i-Martin et al. (2004), Hansen (2007)), as well as regularization methods such as the Lasso (Tibshirani (1996)).

The remainder of the paper is organized as follows. Section 2 introduces the model setup and reviews Leamer's EBA analysis. Section 3 proposes the modified version of EBA and provides the main theoretical results on the asymptotic size of the EBA procedures. Section 4 presents Monte Carlo simulations in various specifications. Section 5 illustrates the proposed methods by revisiting an empirical example of Levine and Renelt (1992). Section 6 concludes. Appendix A provides all the main proofs, and Appendix B describes the variables used in the empirical application.

2 Model Setup and Review of Leamer's EBA

We follow the same model used in Levine and Renelt (1992) and Sala-i-Martin (1997), which is based on versions of EBA from Leamer (1983, 1985) and Leamer and Leonard (1983).

Let $\{(y_i, x_i, w_i, z_i)\}_{i=1}^{n}$ be independent and identically distributed random vectors. The model we consider is

$$y_i = x_i\beta + w_i'\delta + z_i'\gamma + e_i \tag{1}$$

where y_i is a scalar dependent variable, x_i is the scalar variable of interest, w_i is a vector of regressors always included in the regression, z_i is the full set of other available regressors in the data, and e_i is an unobservable error term. Here, (x_i, w_i) are the core regressors, which will be included in all the model specifications that can be based on past empirical studies and economic theory. The variables z_i are auxiliary regressors, which may or may not be included in the model; different subsets are chosen from the pool of z_i in various specifications. If no prior information is available, w_i may include only a constant, and z_i includes all other available variables. $\theta = (\beta, \delta', \gamma')'$ is the unknown parameter vector, and our focus is on the coefficient β associated with the main variable of interest. See Section 5 for how we use this model in the context of growth regressions.

We now formally describe Leamer's EBA procedure.[3] Suppose that we have a set of M sub-

[3]It is important to note that Leamer's original motivation for EBA is based on a Bayesian framework about prior uncertainty in model specification. From a Bayesian standpoint, EBA identifies a plausible range of coefficient estimates, given a prior distribution over possible models. However, in practice, EBA is often implemented and interpreted in a frequentist manner using t-tests for statistical significance across model specifications, as in Levine and Renelt (1992) and Sala-i-Martin (1997). In this paper, we follow the classical hypothesis testing setup and formally analyze the EBA procedure and show that its asymptotic size is much smaller than the desired significance level.

models, where the m-th submodel ($m = 1, ..., M$) includes all core regressors x_i, w_i and a subset of auxiliary regressors $z_{m,i}$ from z_i. The set of models may be nested or non-nested. Throughout the paper, we assume M is fixed.

Let $\widehat{\beta}_m$ denote the estimated OLS regression coefficient for the m-th submodel, with corresponding heteroskedasticity-robust standard error $\widehat{\sigma}_m$. Then, Leamer's EBA is defined as follows for a 5% significance level.

Leamer's EBA: The variable x_i is *"robust"* if the lower and upper extreme bounds have the same sign. Otherwise, the variable is defined as *"fragile (not robust)"*. That is, x_i is *"robust"* if

$$0 \notin [\min_{m=1,...,M}(\widehat{\beta}_m - 1.96\,\widehat{\sigma}_m), \max_{m=1,...,M}(\widehat{\beta}_m + 1.96\,\widehat{\sigma}_m)], \tag{2}$$

and x_i is *"fragile (not robust)"*, otherwise.

When there is no model specification search (i.e., $M = 1$), we observe that determining the variable is "robust" in the EBA is equivalent to rejecting a two-sided t-test for $H_0 : \beta = 0$ (i.e., "significance" of the variable x_i) because

$$|t| = |\widehat{\beta}/\widehat{\sigma}| > 1.96 \Longleftrightarrow 0 \notin [\widehat{\beta} - 1.96\,\widehat{\sigma}, \widehat{\beta} + 1.96\,\widehat{\sigma}].$$

Therefore, in the absence of model specification search, a decision that the variable is "robust" controls the Type-I error probability as usual with the chosen critical value 1.96 for a significance level $\alpha = 0.05$. However, in Leamer's EBA, a variable is "robust" only if it is significantly different from zero for *all* models $m = 1, ..., M$. Through simple calculations, we observe that $\min_m |t_m| > 1.96$ is equivalent to zero not included in the *union* of individual confidence intervals,

$$\min_m |t_m| = \min_m |\widehat{\beta}_m/\widehat{\sigma}_m| > 1.96 \Longleftrightarrow 0 \notin \bigcup_{m=1,...,M} [\widehat{\beta}_m - 1.96\,\widehat{\sigma}_m, \widehat{\beta}_m + 1.96\,\widehat{\sigma}_m].$$

We also observe that

$$0 \notin [\min_m(\widehat{\beta}_m - 1.96\,\widehat{\sigma}_m), \max_m(\widehat{\beta}_m + 1.96\,\widehat{\sigma}_m)] \Longrightarrow \min_m |t_m| > 1.96,$$

since $\bigcup_m[\widehat{\beta}_m - 1.96\widehat{\sigma}_m, \widehat{\beta}_m + 1.96\widehat{\sigma}_m] \subseteq [\min_m(\widehat{\beta}_m - 1.96\widehat{\sigma}_m), \max_m(\widehat{\beta}_m + 1.96\widehat{\sigma}_m)]$, but the converse does not hold in general. In fact, the original Leamer's EBA criterion admits a sharp characterization. Since the extreme bounds interval excludes zero if and only if $\min_m(\widehat{\beta}_m - 1.96\,\widehat{\sigma}_m) > 0$ or $\max_m(\widehat{\beta}_m + 1.96\,\widehat{\sigma}_m) < 0$, we have

$$0 \notin [\min_m(\widehat{\beta}_m - 1.96\,\widehat{\sigma}_m), \max_m(\widehat{\beta}_m + 1.96\,\widehat{\sigma}_m)] \Longleftrightarrow t_m > 1.96 \text{ for all } m, \text{ or } t_m < -1.96 \text{ for all } m.$$

That is, the original EBA criterion requires $\min_m |t_m| > 1.96$ *and* sign agreement of $\widehat{\beta}_m$ across all submodels. The two criteria therefore differ when every individual confidence interval excludes zero, but they do not all lie on the same side. It follows that the original EBA criterion using the extreme bounds interval $[\min_m(\widehat{\beta}_m - 1.96\widehat{\sigma}_m), \max_m(\widehat{\beta}_m + 1.96\widehat{\sigma}_m)]$ is potentially even more conservative than using the minimum of t-statistics directly. We note that EBA is equivalent to the test based on minimum of t-statistics only if we redefine the EBA as the union of CIs, instead of the original extreme bounds as shown above.

Above all, using the standard critical value of 1.96 from a single t-statistic can be too conservative (i.e., it is difficult to reject in all models), because $|t_m| \geq \min_m |t_m|$ for any $m = 1, ..., M$. Many papers in the EBA literature (Sala-i-Martin (1997), Granger and Uhlig (1990)) have criticized the EBA for requiring strong criteria for deciding the robustness of the variables, noting that the results can depend heavily on a single regression model.[4] In our hypothesis testing framework, this critique corresponds to the use of the "minimum" t-statistic, which generally has low power. We point out in this paper that the conservativeness of the EBA also arises from using the standard normal critical value (1.96) from a single t-statistic. If we instead use critical values based on the distribution of the minimum t-statistic (which are no larger than 1.96), the resulting extreme bounds become narrower, which in turn reduces the likelihood that zero falls within the bounds.[5]

3 Modified Extreme Bounds Analysis

We now formally demonstrate the conservativeness of the original EBA procedure and provide modified procedures using new bootstrap critical values based on the minimum t-statistics, which ensures valid asymptotic size.

We note that Leamer's original motivation for EBA is not based on hypothesis testing; rather, it examines the sensitivity of estimates across specifications. However, considering EBA in a classical hypothesis testing framework allows us to formally assess its size properties, which has not been done in the previous literature. Our use of bootstrap critical values for an extremum of studentized statistics across models is related to the data-snooping literature based on the maximum statistic (White (2000), Romano and Wolf (2005)); the EBA criterion instead leads to the minimum of absolute t-statistics, whose null distribution features components that diverge under misspecified submodels.

The EBA procedure evaluates t-statistics across multiple submodels. By the nature of the sensitivity analysis and specification search, we may consider the problem as a multiple testing problem, i.e., $H_0 : \beta_m = 0$ for all $m = 1, \cdots, M$, where β_m denotes the population projection coefficient for the m-th submodel. However, we consider the null hypothesis that the variable x_i

[4] Sala-i-Martin (1997) notes that *"Not surprisingly, Levine and Renelt (1992)'s conclusion is that very few (or no) variables are robust. One possible reason for finding few or no robust variables is, of course, that very few variables can be identified to be correlated systematically with growth....Another explanation, however, is that the test is too strong for any variable to pass it: if the distribution of the estimators of β, has some positive and some negative support, then one is bound to find one regression for which the estimated coefficient changes signs if enough regressions are run. Thus, giving the label of nonrobust to all variables is all but guaranteed."*

[5] At the population level, critical values based on the distribution of the minimum t-statistic are lower than the standard normal critical values. One caveat is that the critical values can be inflated due to the OVB from the misspecified submodels in finite samples, and they can sometimes exceed the standard normal critical values.

has no direct effect on y_i in the true model (1), $H_0 : \beta = 0$; the researcher wants to know whether x_i truly affects y_i, and examines robustness across specifications.

When some auxiliary regressors z_i are correlated with x_i and have nonzero coefficients γ, a submodel that omits these variables will have omitted variable bias (OVB). In order to properly account for OVB across submodels and ensure valid size under $H_0 : \beta = 0$, the bootstrap procedure must be carefully constructed to impose $H_0 : \beta = 0$ while preserving the dependence structure among the regressors. Following the principle that bootstrap resampling should reflect the null hypothesis (Hall and Wilson (1991), Davidson and MacKinnon (1999)), we propose a restricted wild bootstrap procedure. Under $H_0 : \beta = 0$, the true model is $y_i = w_i'\delta + z_i'\gamma + e_i$, which does not include x_i. We estimate this restricted model by OLS to obtain the restricted estimates $\widetilde{\delta}$, $\widetilde{\gamma}$, and the restricted residuals $\widetilde{e}_i = y_i - w_i'\widetilde{\delta} - z_i'\widetilde{\gamma}$. The bootstrap sample is then generated as

$$y_i^* = w_i'\widetilde{\delta} + z_i'\widetilde{\gamma} + \widetilde{e}_i \cdot v_i, \tag{3}$$

where the regressors (x_i, w_i, z_i) are held fixed and v_i are i.i.d. random variables, independent of the data, with $E[v_i] = 0$ and $E[v_i^2] = 1$ (Liu (1988), Mammen (1993)), such as Mammen's (1993) two-point distribution or the Rademacher distribution; see also Flachaire (2005), MacKinnon (2006), and Davidson and Flachaire (2008). Given a bootstrap sample, for each submodel $m = 1, \cdots, M$, we compute $\widehat{\beta}_m^*$ by regressing y_i^* on the corresponding regressors $(x_i, w_i, z_{m,i})$, and the heteroskedasticity-robust standard error $\widehat{\sigma}_m^*$.[6]

An alternative resampling scheme replaces $\widetilde{e}_i \cdot v_i$ in (3) with draws $\widetilde{e}_i^*$ taken randomly with replacement from the centered restricted residuals $\{\widetilde{e}_i - \bar{\widetilde{e}}\}_{i=1}^n$, i.e., an i.i.d. restricted residual bootstrap. As discussed in Remark 1 below, this alternative is asymptotically valid under conditional homoskedasticity, but not under general heteroskedasticity, which is why we choose the wild bootstrap in (3) as the default choice.

The key feature of this restricted bootstrap is that the bootstrap data are generated under $H_0 : \beta = 0$, and the OVB structure is preserved; submodels that omit auxiliary regressors with nonzero $\widetilde{\gamma}$ will exhibit the same OVB pattern in the bootstrap world as in the original sample. In contrast, a recentered nonparametric bootstrap that uses $t_m^* = (\widehat{\beta}_m^* - \widehat{\beta}_m)/\widehat{\sigma}_m^*$ eliminates this OVB from the bootstrap distribution, which can lead to size distortions when the OVB is non-negligible. In our simulations, we compare the finite-sample performance of these alternative methods.

We define the bootstrap critical value $\widehat{c}_{1-\alpha}$ as the $(1-\alpha)\%$ quantile of the minimum of the absolute bootstrap t-statistics, and the bootstrap critical value is computed by the following steps:

(1) Estimate the restricted model by regressing y_i on (w_i, z_i) (excluding x_i) by OLS, and obtain the restricted estimates $\widetilde{\delta}$, $\widetilde{\gamma}$ and the restricted residuals $\widetilde{e}_i = y_i - w_i'\widetilde{\delta} - z_i'\widetilde{\gamma}$, $i = 1, \ldots, n$.

(2) For each bootstrap replication $b = 1, \ldots, B$, draw $\{v_i(b)\}_{i=1}^n$ i.i.d. from the chosen weight

[6] See Freedman (1981) for bootstrap consistency in linear regression, and MacKinnon (2006) for the use of restricted and unrestricted models in bootstrap hypothesis testing.

distribution, and generate the bootstrap sample $y_i^*(b) = w_i'\widetilde{\delta} + z_i'\widetilde{\gamma} + \widetilde{e}_i \cdot v_i(b)$ as in (3), holding the regressors (x_i, w_i, z_i) fixed.

(3) For each submodel $m = 1, \ldots, M$, regress $y_i^*(b)$ on $(x_i, w_i', z_{m,i}')'$ by OLS to obtain $\widehat{\beta}_m^*(b)$ and its heteroskedasticity-robust standard error $\widehat{\sigma}_m^*(b)$, and compute the minimum absolute bootstrap t-statistic $\min_{m=1,\ldots,M} |\widehat{\beta}_m^*(b)/\widehat{\sigma}_m^*(b)|$.

(4) Set the bootstrap critical value $\widehat{c}_{1-\alpha}$ as follows;

$$\widehat{c}_{1-\alpha} \equiv (1-\alpha)\% \text{ conditional quantile of } \{\min_{m=1,\ldots,M} |\widehat{\beta}_m^*(b)/\widehat{\sigma}_m^*(b)|\}_{b=1}^B \text{ given the data.} \quad (4)$$

Then, we define the modified Leamer's EBA with the bootstrap critical value $\widehat{c}_{1-\alpha}$ instead of normal critical values as follows:

Modified Leamer's EBA: The variable x_i is *"robust"* if

$$\min_{m=1,\ldots,M} |\widehat{\beta}_m/\widehat{\sigma}_m| > \widehat{c}_{1-\alpha}, \quad (5)$$

and x_i is *"fragile (not robust)"* otherwise.[7]

We now provide the main theoretical results. We consider the case where the auxiliary regression coefficients γ in the full model (1) are fixed (do not vary with n). In this case, the OVB for misspecified submodels will dominate, so the t-statistics from misspecified models diverge as $n \to \infty$. Consequently, the minimum t-statistic is asymptotically determined by the correctly specified models.

To state the main result formally, we introduce further notation and regularity conditions. For each submodel m, let $\widetilde{x}_{m,i}$ denote the residual from the population linear projection of x_i onto $(w_i', z_{m,i}')'$, let $z_{-m,i}$ denote the auxiliary regressors omitted from submodel m and γ_{-m} the corresponding subvector of γ, and define $Q_m \equiv E[\widetilde{x}_{m,i}^2]$ and $R_m \equiv E[\widetilde{x}_{m,i} z_{-m,i}']$, so that the population projection coefficient of submodel m satisfies $\beta_m = Q_m^{-1} R_m \gamma_{-m}$ under $H_0 : \beta = 0$.

Assumption 1.

(i) $\{(y_i, x_i, w_i, z_i)\}_{i=1}^n$ *are i.i.d. random vectors generated by model (1) with* $E[e_i \mid x_i, w_i, z_i] = 0$. $E[\|q_i\|^4] < \infty$, $E[e_i^4] < \infty$, *and* $\lambda_{\min}(E[q_i q_i']) \geq \underline{\lambda}$ *for some constant* $\underline{\lambda} > 0$, *where* $q_i \equiv (x_i, w_i', z_i')'$, $\lambda_{\min}(A)$ *denotes the smallest eigenvalue of a symmetric matrix* A, *and* $\|\cdot\|$ *denotes the Euclidean norm.* $E[e_i^2 \mid x_i, w_i, z_i] \geq \underline{\sigma}^2$ *almost surely for some constant* $\underline{\sigma}^2 > 0$.

[7] As in Section 2, $0 \notin [\min_m(\widehat{\beta}_m - \widehat{c}_{1-\alpha}\widehat{\sigma}_m), \max_m(\widehat{\beta}_m + \widehat{c}_{1-\alpha}\widehat{\sigma}_m)]$ implies $\min_m |\widehat{\beta}_m/\widehat{\sigma}_m| > \widehat{c}_{1-\alpha}$ for any $\widehat{c}_{1-\alpha} > 0$, but the converse does not hold in general. Hence, defining the modified EBA in terms of extreme bounds intervals as in the original EBA would yield a strictly more conservative criterion than (5), and we therefore use (5) as the preferred form of the modified Leamer's EBA.

(ii) At least one submodel $m^* \in \{1, 2, \cdots, M\}$ satisfies $\gamma_{-m^*} = 0$. Moreover, $R_m \gamma_{-m} \neq 0$ for all $m = 1, ..., M$ with $\gamma_{-m} \neq 0$.

(iii) The wild bootstrap weights $\{v_i\}_{i=1}^n$ in (3) are i.i.d., independent of the data, with $E[v_i] = 0$, $E[v_i^2] = 1$, and $E[v_i^4] < \infty$.

Assumption 1(i) imposes a correct specification assumption for the full model (1) with standard moment and nondegeneracy conditions. The conditional mean restriction $E[e_i \mid x_i, w_i, z_i] = 0$ is stronger than the linear projection, but under this assumption, the restricted bootstrap mimics the joint sampling distribution of the t-statistics across submodels. The eigenvalue bound rules out perfect multicollinearity and implies $\min_{1 \leq m \leq M} Q_m \geq \underline{\lambda} > 0$.

The first part of Assumption 1(ii) requires at least one correctly specified submodel among the M candidates, as discussed above. We say that a submodel m^* is *correctly specified* if its population projection coefficient satisfies $\beta_{m^*} = 0$ under H_0. This holds whenever submodel m^* includes all auxiliary regressors z_j with nonzero coefficients γ_j; it need not include all regressors in z_i, and in particular it need not be the full model. Including additional regressors with zero coefficients also yields a correctly specified model, though possibly with efficiency loss.

Under the second part of Assumption 1(ii), the set of correctly specified submodels satisfies $\mathcal{S}_0 \equiv \{m : \beta_m = 0\} = \{m : \gamma_{-m} = 0\}$, and every submodel in $\mathcal{S}_0$ has population projection error equal to the true error e_i. Assumption 1(iii) is satisfied by Mammen's (1993) two-point distribution and by Rademacher weights (Davidson and Flachaire (2008)).

Theorem 1. *Suppose Assumption 1 holds and let $\alpha \in (0, 1/2)$. Then, under $H_0 : \beta = 0$ with fixed γ, we have*

$$\lim_{n\to\infty} P(\min_{m=1,\ldots,M} |\widehat{\beta}_m / \widehat{\sigma}_m| > z_{1-\alpha/2}) \leq \alpha, \tag{6}$$

$$\lim_{n\to\infty} P(\min_{m=1,\ldots,M} |\widehat{\beta}_m / \widehat{\sigma}_m| > \widehat{c}_{1-\alpha}) = \alpha, \tag{7}$$

where $z_{1-\alpha/2}$ is the $(1 - \alpha/2)$ quantile of the standard normal distribution, and the critical value $\widehat{c}_{1-\alpha}$ is defined in (4) using the restricted wild bootstrap (3). Moreover, the original Leamer's EBA based on the extreme bounds interval is at least as conservative as the $\min_m |t_m|$ test in (6), since $0 \notin [\min_m(\widehat{\beta}_m - z_{1-\alpha/2}\widehat{\sigma}_m), \max_m(\widehat{\beta}_m + z_{1-\alpha/2}\widehat{\sigma}_m)]$ implies $\min_m |\widehat{\beta}_m / \widehat{\sigma}_m| > z_{1-\alpha/2}$.

The conservativeness of Leamer's EBA in (6) follows from the fact that $\min_m |t_m| \leq |t_{m^*}|$, and the t-statistic from the correctly specified model m^* has the usual standard normal limiting distribution since $\beta_{m^*} = 0$. For the equality in (7), under fixed γ, all misspecified models have $|t_m| \to \infty$, so the minimum is asymptotically determined by the correctly specified models. The restricted wild bootstrap correctly reproduces both the divergent behavior of misspecified models and the joint centered Gaussian behavior of correctly specified models, which ensures that the bootstrap critical value converges to the correct quantile of the limiting distribution of $\min_m |t_m|$ under H_0.

Remark 1 (Wild bootstrap versus i.i.d. residual bootstrap under heteroskedasticity)**.** Theorem 1 employs the wild bootstrap in (3). The limiting null distribution of $\min_m |t_m|$ is the minimum of $|Z_m|$, $m \in \mathcal{S}_0$, and hence depends on the entire correlation matrix of a mean-zero Gaussian vector with $E[Z_j Z_l] = \Omega_{jl}/(\Omega_{jj}^{1/2}\Omega_{ll}^{1/2})$, where $\Omega_{jl} \equiv E[\widetilde{x}_{j,i}\widetilde{x}_{l,i}\, e_i^2]$ for $j, l \in \mathcal{S}_0$. Conditional on the data, the wild bootstrap scores have covariance $n^{-1}\sum_i \widehat{\widetilde{x}}_{j,i}\widehat{\widetilde{x}}_{l,i}\widetilde{e}_i^2 \xrightarrow{p} \Omega_{jl}$, where $\widehat{\widetilde{x}}_{j,i}$ denotes the residual from the sample regression of x_i on $(w_i', z_{j,i}')'$ (the sample counterpart of $\widetilde{x}_{j,i}$), so the cross-model correlation structure is replicated. In contrast, i.i.d. resampling from the centered restricted residuals yields conditional covariance $\widehat{s}^2\, n^{-1}\sum_i \widehat{\widetilde{x}}_{j,i}\widehat{\widetilde{x}}_{l,i} \xrightarrow{p} E[e_i^2]\, E[\widetilde{x}_{j,i}\widetilde{x}_{l,i}]$, where $\widehat{s}^2$ is the sample variance of the centered restricted residuals. The two limits coincide under conditional homoskedasticity, $E[e_i^2 \mid x_i, w_i, z_i] = \sigma^2$, but differ in general.

Remark 2 (Divergent t-statistics and the continuous mapping theorem)**.** Under fixed γ, the joint vector $(t_1, \ldots, t_M)'$ does not converge in distribution to a bounded random vector: the components corresponding to misspecified submodels diverge to $\pm\infty$. Thus, the standard continuous mapping theorem cannot be applied directly. In the proof, we instead apply a continuous mapping argument on the extended Euclidean space $\mathbb{R}\cup\{\pm\infty\}$, for the minimum functional (Lemma 1 in Appendix A), adapting a similar device from the moment inequality literature, e.g., Andrews and Guggenberger (2009).

4 Simulation Study

We investigate the finite sample performance of Leamer's EBA using Monte Carlo experiments. In particular, we compare the finite sample size of EBA using the proposed bootstrap critical values against the original EBA.

We consider the following linear regression model

$$y_i = \delta + \beta x_i + \sum_{j=1}^{K} \gamma_j z_{ji} + e_i$$

where $(x_i, z_{1i}, ..., z_{Ki})' \sim N(0, \Sigma)$. The diagonal elements of Σ are 1 and the off-diagonal elements are ρ. We set $\rho = 0.5$. The error term follows $e_i \sim N(0, \sigma_i^2)$, with $\sigma_i = 2$ for the homoskedastic setup, and $\sigma_i = |1 + 2x_i|/3$ for the heteroskedastic case. We include a constant and x_i as the core regressors in every model, and consider all z_i as auxiliary regressors. For the sample sizes (n) and the maximum number of auxiliary regressors (K), we consider $(n, K) \in \{(100, 4), (200, 4), (500, 6)\}$. We consider all possible combinations of submodels, resulting in a total of $M = 2^K$ models. For the regression coefficients, we set $\beta = 0$, $\delta = 1$ and we consider $\gamma = (c, c^2, c^3, c^4)'$ for $K = 4$ and $\gamma = (c, c^2, c^3, c^4, 0, 0)'$ for $K = 6$. We consider $c \in \{0.1, 0.5, 0.9\}$. Since x_i is correlated with z_i and $\gamma \neq 0$, any submodel that omits auxiliary regressors is misspecified and exhibits omitted variable bias (OVB). The scalar c controls the degree of misspecification and magnitude of the OVB: $c = 0.1$ corresponds to the small OVB, while $c = 0.9$ leads to substantial OVB in misspecified submodels. The number of MC replications is 1,000 and the bootstrap replications $B = 999$. We also considered $\rho \in \{0.2, 0.8\}$, but the results are qualitatively similar, so they are not reported here for brevity.

Table 1 reports the empirical size of Leamer's EBA at nominal $\alpha = 5\%$, i.e., the probability of (incorrectly) declaring x_i "robust" when $\beta = 0$. We compute this as $P(\min_m |\widehat{\beta}_m/\widehat{\sigma}_m| > c)$, where the critical value c is either 1.96 (original EBA) or the bootstrap critical value $\widehat{c}_{0.95}$ (modified EBA). We compare three bootstrap variants for $\widehat{c}_{0.95}$. The first is a naive nonparametric bootstrap (NB) that resamples the original observations with replacement and uses recentered t-statistics, $\min_m |(\widehat{\beta}^*_m - \widehat{\beta}_m)/\widehat{\sigma}_m|$. The other two (the restricted residual bootstrap (RB) and the wild bootstrap (WB)) impose $H_0 : \beta = 0$ using the restricted residuals defined in Section 3: RB resamples them with replacement, while WB uses the weights drawn from Mammen's (1993) two-point distribution. As a benchmark, we also report the infeasible 95% critical value from the simulated distribution of $\min_m |t_m|$, denoted cv (min-$|t|$).

Table 1 delivers three main findings. First, the original Leamer's EBA is severely conservative across all designs: empirical sizes range from 0.011 to 0.030, well below the nominal 5% level, and the under-rejection is essentially insensitive to the DGP. This is consistent with Theorem 1: the standard normal cutoff of 1.96 is too strict for the minimum.

The naive recentered bootstrap (NB) fails to control size when OVB is non-negligible, and this worsens with increasing c. For $c = 0.1$, NB sizes remain moderate (0.036-0.069), but at $c = 0.5$, the size distortion is noticeable (0.071-0.108), and at $c = 0.9$, the size distortion is severe (0.137-0.165 in the homoskedastic design and 0.253-0.310 under heteroskedasticity). The reason is straightforward: NB recenters the bootstrap t-statistics at $\widehat{\beta}_m$, which eliminates the OVB from the bootstrap distribution. As c increases, the gap between the bootstrap distribution and the true null distribution of $\min_m |t_m|$ widens, and NB performs poorly.

Finally, the restricted residual methods (RB and WB) control size close to the nominal 5% across designs. Both procedures exhibit mild over-rejection at $c = 0.1$, but their sizes are closer to the nominal level for moderate and large c. The over-rejection of RB at $c = 0.1$ (0.095-0.103 under heteroskedasticity) persists even at $n = 500$. The distortion is largest when the OVB is small, since then many submodels are nearly correctly specified and the null distribution of $\min_m |t_m|$ depends heavily on the entire correlation matrix of the t-statistics, which the i.i.d. resampling scheme approximates less accurately in finite samples than the wild bootstrap under heteroskedasticity (see Remark 1). WB is comparable to RB under homoskedasticity and performs well under heteroskedasticity, making it a reasonable default choice.

The infeasible benchmark cv (min-$|t|$) provides some further insights. Across all designs it lies between 1.16 and 1.69, which is uniformly below 1.96, and this is exactly why the original EBA is conservative. The bootstrap critical values from RB and WB track this benchmark closely, although not reported here for brevity, as reflected in their empirical sizes. In sum, the simulation evidence confirms Theorem 1 and supports the modified EBA with restricted residual bootstrap

Table 1: Empirical size of the EBA procedures at the nominal level $\alpha = 5\%$ under $\beta = 0$.

	Homoskedastic			Heteroskedastic		
c	0.1	0.5	0.9	0.1	0.5	0.9
$n = 100, K = 4$						
Leamer's EBA	0.013	0.018	0.027	0.019	0.027	0.023
Modified EBA (NB)	0.064	0.073	0.137	0.069	0.102	0.253
Modified EBA (RB)	0.072	0.057	0.050	0.096	0.057	0.063
Modified EBA (WB)	0.076	0.066	0.059	0.074	0.079	0.076
cv (min-$\|t\|$)	1.39	1.47	1.66	1.64	1.69	1.68
$n = 200, K = 4$						
Leamer's EBA	0.013	0.022	0.026	0.023	0.030	0.030
Modified EBA (NB)	0.046	0.071	0.145	0.062	0.108	0.273
Modified EBA (RB)	0.051	0.047	0.043	0.103	0.062	0.057
Modified EBA (WB)	0.056	0.052	0.042	0.069	0.069	0.069
cv (min-$\|t\|$)	1.27	1.47	1.55	1.61	1.69	1.69
$n = 500, K = 6$						
Leamer's EBA	0.012	0.021	0.026	0.011	0.017	0.017
Modified EBA (NB)	0.056	0.096	0.165	0.036	0.103	0.310
Modified EBA (RB)	0.072	0.065	0.058	0.095	0.060	0.057
Modified EBA (WB)	0.070	0.070	0.062	0.054	0.055	0.052
cv (min-$\|t\|$)	1.23	1.51	1.16	1.30	1.48	1.49

Notes: (1) Leamer's EBA denotes the original Leamer's Extreme Bound Analysis using the standard normal critical values. We consider two different bootstrap critical values for the Modified EBA based on restricted residuals; 1) RB - residual bootstrap, 2) WB - wild bootstrap with the Mammen's (1993) two-point distribution. We also consider critical values from the Naive Bootstrap with recentered t-statistics (NB). We denote cv (min-$|t|$) as (infeasible) 95% critical values from distributions of minimum absolute t-statistics. (2) We calculate the size of the EBA procedures as follows; $P(\min_m |\widehat{\beta}_m/\widehat{\sigma}_m| > 1.96)$ and $P(\min_m |\widehat{\beta}_m/\widehat{\sigma}_m| > \widehat{c}_{0.95})$, i.e., the probability that the variable x_i is considered "robust" in the EBA procedures when $\beta = 0$. (3) n is the sample size, K is the maximum number of auxiliary regressors.

critical values, with WB as a convenient default.

5 Empirical Application: Levine and Renelt (1992)

In this section, we illustrate our proposed methods by revisiting Levine and Renelt (1992, hereafter LR92). LR92 considers specification analysis in the empirical cross-country growth regression literature. Specifically, they consider the following model, the same as in equation (1):

$$y_i = x_i\beta + w_i'\delta + z_i'\gamma + e_i,$$

where y_i is the average annual growth rate of GDP per capita (GDP), and w_i is chosen based on past empirical studies, typically including the investment share of GDP (INV), the initial level of real GDP per capita (RGDP60), and the initial secondary school enrollment rate (SEC).

The variables z_i can include all other explanatory variables whose inclusion in the regression is uncertain. In practice, the number of doubtful variables $z_{m,i}$ can be limited, so the total number of models M is also limited. See Appendix B and also LR92 for potential z variables for the specification search, and for more detailed descriptions.[8]

Table 2 replicates the results in Levine and Renelt (1992, Tables 1, 6, 8, 10), when the dependent variable is the growth rate of real per capita GDP (GDP). Note that the variable of interest ("x" variable in our setup, and "M"-variable in LR92 notation) changes and we follow the same approach as in Levine and Renelt (1992) by limiting the number of doubtful variables per model to at most 3 ("LR92 - Restricted") and use the same "z" ("other variable" in LR92 notation) variables as in LR92. Depending on the "group" of variable x, we report sensitivity results for Core Variables (LR92, Table 1), Fiscal Variables (LR92, Table 6), Trade Variables (LR92, Table 8), and Monetary and Political Variables (LR92, Table 10) altogether in Table 2.

We report the minimum of absolute t-statistics (min-$|t|$) with the proposed critical values (cv) based on the restricted wild bootstrap defined in (4). In all cases, critical values are calculated based on $B = 9,999$ bootstrap samples. Our main emphasis is on comparing the original EBA procedures to the modified procedure replicating LR92; however, we also investigate relaxed model consideration sets by allowing a larger number of doubtful variables per model up to 7 ("Unrestricted"), which yields a total of $M = 2^7$ models.

In Table 2, variables are categorized as "Kept Robust" if they are classified as robust both in the original and the modified procedures (i.e., min-$|t| > 1.96$ and min-$|t| > cv$). Variables are categorized as "Turned Robust" if they were classified as fragile in LR92 with the original EBA, but

[8] LR92 uses the combined datasets from the Penn World Tables as well as IMF and World Bank resources. The data cover the 1960-1989 time period, while the accompanying dataset, spanning from 1974 to 1989, covers a more extensive set of variables for this shorter time frame and enriches the former with additional fiscal policy and fiscal monitoring variables. The dataset includes 119 countries, but primarily focuses on 109 non-Oil exporter/OPEC countries. The original datasets are assembled in the online data appendix of Ericsson et al. (2001) in the Journal of Applied Econometrics database and the main dataset can be found in: https://journaldata.zbw.eu/dataset/output-and-inflation-in-the-long-run.

classified as robust with the modified procedure (i.e., min-$|t| < 1.96$, but min-$|t| > cv$). Otherwise, variables are categorized as "Kept Fragile".[9]

Interestingly, many variables categorized as "fragile (not robust)" in LR92 are now classified as "robust" with the modified EBA procedures. Specifically, one of the "core" variables, the population growth (GPO) became robust, although it was originally labelled as "fragile" in LR92. Additionally, government consumption less defense and education share of GDP (GOVX), and one of the Trade variables (RERDB) became "robust" with the modified critical values compared with LR92. For most of these variables, however, their min-$|t|$ statistics only marginally exceed the critical value.

As expected, many of the core variables, including investment share (INV), the initial per capita GDP (RGDP60), and the secondary school enrollment rate (SEC), remain "robust", and many other variables still remain "fragile". Increasing the total number of models ("Unrestricted") makes both min-$|t|$ statistics and critical values weakly smaller, thus results are generally similar to the original LR92 case (restricted), except for the GOVX variable, which is classified as "robust" in the restricted case, but "fragile" in the unrestricted case.

In addition, Table 3 replicates Levine and Renelt (1992, Tables 2, 7, 9, 11), using the investment share (INV) as the dependent variable. As in LR92, we use the same set of x variables ("variables of interest") as well as z variables ("other variables"), and no w variables are included. In Table 3, we find that no variable changes its classification under the modified procedure.

6 Conclusion

In this paper, we revisit Leamer's EBA within a classical hypothesis testing framework. We first show that the original EBA procedure is asymptotically conservative, and consequently classifies too many variables as "fragile" ("not robust"). To address this issue, we propose a modified EBA procedure that utilizes bootstrap critical values from the distribution of minimum t-statistics, implemented by a restricted wild bootstrap that imposes the null hypothesis while preserving the omitted variable bias structure across submodels. We establish the asymptotic validity of this procedure under fixed auxiliary regression coefficients γ (Theorem 1). Simulation evidence supports the theoretical findings, and we illustrate our proposed method in an empirical application of Levine and Renelt (1992), and find that the modified procedure reclassifies some variables previously deemed "fragile" as "robust".

[9] There are some minor differences between Table 2 and the Tables from LR92. First of all, LR92 uses their original criterion for classification of "robust" variables, i.e., $0 \notin [\min_m(\widehat{\beta}_m - 1.96\,\widehat{\sigma}_m), \max_m(\widehat{\beta}_m + 1.96\,\widehat{\sigma}_m)]$, which we show in this paper is more conservative than using the minimum of t-statistics directly. Furthermore, we noticed that t-statistics in LR92 are based on the homoskedastic version of standard errors while we use the heteroskedasticity-robust standard errors, which can sometimes be twice as large as homoskedastic standard errors, and this makes min-$|t|$ smaller. However, the vast majority of results are qualitatively similar to the LR92.

Table 2: Modified EBA Results: dependent variable (GDP)

x	Group	LR92-Restricted min-$\|t\|$	LR92-Restricted cv	Unrestricted min-$\|t\|$	Unrestricted cv
			Kept Robust		
INV	Core	3.968	1.155	3.963	1.131
RGDP60	Core	2.482	2.357	2.264	1.743
SEC	Core	2.136	1.702	1.877	1.492
STDD	Monetary	2.136	1.813	2.136	1.810
			Turned Robust		
GPO	Core	1.580	1.546	1.578	1.522
GOVX	Fiscal	1.888	1.756	1.432	1.507 (F)
RERDB	Trade	1.556	1.544	1.345	1.135
PI	Monetary	0.935	0.931	0.935	0.931
			Kept Fragile		
GOV	Fiscal	0.239	0.914	0.239	0.910
TEX	Fiscal	0.414	0.575	0.392	0.546
DEF	Fiscal	0.582	1.080	0.568	1.076
EXP	Trade	0.111	1.460	0.111	1.395
IMP	Trade	0.060	1.153	0.026	1.077
LEAM1	Trade	0.006	1.069	0.006	0.882
LEAM2	Trade	0.108	1.044	0.108	0.910
BMP	Trade	1.027	1.466	0.304	0.945
STDI	Monetary	0.803	0.891	0.803	0.891
GDC	Monetary	0.020	2.922	0.020	2.852
REVC	Political	0.194	1.370	0.145	1.223

Note. This table replicates Levine and Renelt (1992, Tables 1, 6, 8, 10). "LR92 - Restricted" uses the same approach as in LR 92, limiting the number of doubtful variables per model to 3. "Unrestricted" uses a larger number of doubtful variables per model up to 7 (total number of models $M = 2^7$). We report the minimum of absolute t-statistics (min-$|t|$) with the proposed critical values (cv) based on wild bootstrap.

Variables are labeled as "Kept Robust" if they are classified as robust both in the original and the modified procedures (i.e., min-$|t| > 1.96$ and min-$|t| > cv$). Variables are categorized as "Turned Robust" if they were classified as fragile in LR92 with the original EBA, but classified as robust with the modified procedure (i.e., min-$|t| < 1.96$, but min-$|t| > cv$). Otherwise, variables are categorized as "Kept Fragile". These classifications are based on the LR92-Restricted specification. The marker "(F)" indicates that the variable is classified as fragile (min-$|t| < cv$) under the corresponding specification.

Table 3: Modified EBA Results: dependent variable (INV)

x	Group	LR92-Restricted min-$\|t\|$	LR92-Restricted cv	Unrestricted min-$\|t\|$	Unrestricted cv
Kept Robust					
EXP	Trade	3.734	2.116	3.633	1.871
LEAM2	Trade	2.925	2.110	2.905	1.895
REVC	Political	2.075	1.083	2.075	1.064
BMP	Trade	2.792	2.110	2.539	2.019
Turned Robust					
N/A					
Kept Fragile					
RGDP60	Core	0.022	0.729	0.022	0.713
GPO	Core	0.360	1.431	0.157	1.332
SEC	Core	0.751	1.225	0.751	1.204
GOV	Fiscal	0.821	1.460	0.732	1.418
GOVX	Fiscal	0.031	1.056	0.031	0.925
TEX	Fiscal	1.110	1.831	0.974	1.566
DEF	Fiscal	0.018	1.501	0.018	1.348
RERDB	Trade	0.732	1.307	0.732	1.287
GDC	Monetary	0.457	1.052	0.457	1.052
PI	Monetary	0.032	0.203	0.032	0.202
STDI	Monetary	0.060	0.505	0.060	0.500
STDD	Monetary	0.729	1.564	0.692	1.515
LEAM1	Trade	1.121	1.579	1.121	1.527

Note. This table replicates Levine and Renelt (1992, Tables 2, 7, 9, 11). "LR92 - Restricted" uses the same approach as in LR 92, limiting the number of doubtful variables per model to 3. "Unrestricted" uses a larger number of doubtful variables per model up to 7 (total number of models $M = 2^7$). We report the minimum of absolute t-statistics (min-$|t|$) with the proposed critical values (cv) based on wild bootstrap.
Variables are labeled as "Kept Robust" if they are classified as robust both in the original and the modified procedures (i.e., min-$|t| > 1.96$ and min-$|t| > cv$). Variables are categorized as "Turned Robust" if they were classified as fragile in LR92 with the original EBA, but classified as robust with the modified procedure (i.e., min-$|t| < 1.96$, but min-$|t| > cv$). Otherwise, variables are categorized as "Kept Fragile". These classifications are based on the LR92-Restricted specification.

References

Andrews, D. W. K., and P. Guggenberger. (2009) "Validity of Subsampling and 'Plug-in Asymptotic' Inference for Parameters Defined by Moment Inequalities." *Econometric Theory*, 25(3), 669-709.

Barro, R. J. (1991) "Economic Growth in a Cross Section of Countries." *Quarterly Journal of Economics*, 106(2), 407-443.

Bartley, W. A., and M. A. Cohen. (1998) "The Effect of Concealed Weapons Laws: An Extreme Bound Analysis." *Economic Inquiry*, 36(2), 258-265.

Cameron, A. C., J. B. Gelbach, and D. L. Miller. (2008) "Bootstrap-Based Improvements for Inference with Clustered Errors." *The Review of Economics and Statistics*, 90(3), 414-427.

Davidson, R., and E. Flachaire. (2008) "The Wild Bootstrap, Tamed at Last." *Journal of Econometrics*, 146(1), 162-169.

Davidson, R., and J. G. MacKinnon. (1999) "The Size Distortion of Bootstrap Tests." *Econometric Theory*, 15(3), 361-376.

Davidson, R., and J. G. MacKinnon. (2010) "Wild Bootstrap Tests for IV Regression." *Journal of Business & Economic Statistics*, 28(1), 128-144.

Djogbenou, A. A., J. G. MacKinnon, and M. Ø. Nielsen. (2019) "Asymptotic Theory and Wild Bootstrap Inference with Clustered Errors." *Journal of Econometrics*, 212(2), 393-412.

Dollar, D. (1992) "Outward-Oriented Developing Economies Really Do Grow More Rapidly: Evidence from 95 LDCs, 1976-1985." *Economic Development and Cultural Change*, 40(3), 523-544.

Ericsson, N., J. Irons, and R. Tryon. (2001) "Output And Inflation In The Long Run." *Journal of Applied Econometrics*, 16(3), 241-253.

Flachaire, E. (2005) "Bootstrapping Heteroskedastic Regression Models: Wild Bootstrap vs. Pairs Bootstrap." *Computational Statistics and Data Analysis*, 49(2), 361-376.

Fowles, R., and M. Merva. (1996) "Wage Inequality and Criminal Activity: An Extreme Bounds Analysis for the United States, 1975-1990." *Criminology*, 34(2), 163-182.

Freedman, D. A. (1981) "Bootstrapping Regression Models." *The Annals of Statistics*, 9(6), 1218-1228.

Granger, C. W. J., and H. F. Uhlig. (1990) "Reasonable Extreme-Bounds Analysis." *Journal of Econometrics*, 44(1-2), 159-170.

Hall, P., and S. R. Wilson. (1991) "Two Guidelines for Bootstrap Hypothesis Testing." *Biometrics*, 47, 757-762.

Hansen, B. E. (2007) "Least Squares Model Averaging." *Econometrica*, 75(4), 1175-1189.

Hjort, N. L., and G. Claeskens. (2003) "Frequentist Model Average Estimators." *Journal of the American Statistical Association*, 98(464), 879-899.

Hlavac, M. (2016) "ExtremeBounds: Extreme Bounds Analysis in R." *Journal of Statistical Software*, 72(9), 1-22.

Horowitz, J. L. (2001) "The Bootstrap." In *Handbook of Econometrics*, Vol. 5, ed. by J. J. Heckman and E. E. Leamer, 3159-3228. Elsevier, Amsterdam.

Impavido, G. (1998) "EBA: Stata Module to Perform Extreme Bound Analysis." Statistical Software Components. http://fmwww.bc.edu/repec/bocode/e/eba.ado.

Leamer, E. E. (1983) "Let's Take the Con Out of Econometrics." *The American Economic Review*, 73(1), 31-43.

Leamer, E. E. (1985) "Sensitivity Analyses Would Help." *The American Economic Review*, 75(3), 308-313.

Leamer, E. E. (1988) "Measures of Openness." In *Trade Policy Issues and Empirical Analysis*, ed. by R. E. Baldwin, National Bureau of Economic Research Conference Report Series, Chicago: University of Chicago Press, 147-200.

Leamer, E. E., and H. Leonard. (1983) "Reporting the Fragility of Regression Estimates." *The Review of Economics and Statistics*, 65(2), 306-317.

Levine, R., and D. Renelt. (1992) "A Sensitivity Analysis of Cross-Country Growth Regressions." *The American Economic Review*, 82(4), 942-963.

Levine, R., and S. Zervos. (1993) "What We Have Learned About Policy and Growth from Cross-Country Regressions?" *The American Economic Review*, 83(2), 426-430.

Liu, R. Y. (1988) "Bootstrap Procedures under some Non-I.I.D. Models." *The Annals of Statistics*, 16(4), 1696-1708.

MacKinnon, J. G. (2006) "Bootstrap Methods in Econometrics." *Economic Record*, 82(1), 2-18.

MacKinnon, J. G. (2009) "Bootstrap Hypothesis Testing." In *Handbook of Computational Econometrics*, ed. by D. A. Belsley and E. J. Kontoghiorghes, 183-213. Wiley, Chichester.

MacKinnon, J. G., M. Ø. Nielsen, and M. D. Webb. (2023) "Cluster-Robust Inference: A Guide to Empirical Practice." *Journal of Econometrics*, 232(2), 272-299.

Mammen, E. (1993) "Bootstrap and Wild Bootstrap for High Dimensional Linear Models." *The Annals of Statistics*, 21(1), 255-285.

Moser, C., and J. E. Sturm. (2011) "Explaining IMF Lending Decisions After the Cold War." *Review of International Organizations*, 6, 307-340.

Romano, J. P., and M. Wolf. (2005) "Stepwise Multiple Testing as Formalized Data Snooping." *Econometrica*, 73(4), 1237-1282.

Sala-i-Martin, X. (1997) "I Just Ran Two Million Regressions." *The American Economic Review*, 87(2), 178-183.

Sala-i-Martin, X., G. Doppelhofer, and R. I. Miller. (2004) "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach." *The American Economic Review*, 94(4), 813-835.

Sturm, J. E., and J. de Haan. (2005) "Determinants of Long-Term Growth: New Results Applying Robust Estimation and Extreme Bounds Analysis." *Empirical Economics*, 30(3), 597-617.

Tibshirani, R. (1996) "Regression Shrinkage and Selection Via the Lasso." *Journal of the Royal Statistical Society, Series B (Methodological)*, 58(1), 267-288.

van der Vaart, A. W. (1998) *Asymptotic Statistics.* Cambridge University Press, Cambridge.

White, H. (2000) "A Reality Check for Data Snooping." *Econometrica*, 68(5), 1097-1126.

Wu, C. F. J. (1986) "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis." *The Annals of Statistics*, 14(4), 1261-1295.

Yang, H., J. He, and S. Chen. (2015) "The Fragility of the Environmental Kuznets Curve: Revisiting the hypothesis with Chinese data via an Extreme Bound Analysis." *Ecological Economics*, 109, 41-58.

A Proof of the Main Results

Let the superscript * denote a probability computed under the bootstrap distribution conditional on the original data. Let $\xrightarrow{p^*}$ and $\xrightarrow{d^*}$ denote convergence in probability and convergence in distribution, respectively, under the bootstrap measure conditional on the data, where both are understood to hold in probability with respect to the sampling distribution of the data. We write $X_n^* = O_{p^*}(a_n)$ if, for every $\epsilon > 0$, there exists a constant $M_\epsilon < \infty$ such that $P^*(|X_n^*| > M_\epsilon a_n) \leq \epsilon$ with probability approaching one, and $X_n^* = o_{p^*}(a_n)$ if $X_n^*/a_n \xrightarrow{p^*} 0$. For nonnegative sequences, $a \lesssim b$ means $a \leq Cb$ for some finite constant C that does not depend on n.

The following lemma is an extended continuous mapping result for the minimum of absolute t-statistics when some components diverge. It adapts the transformation device of Andrews and Guggenberger (2009) from the moment inequality literature, which allows some components of the vector to diverge in probability.

Lemma 1. *Let $T_n = (T_{n,1}, \ldots, T_{n,M})'$ be random vectors in $\mathbb{R}^M$, and let $\{1, \ldots, M\} = \mathcal{S}_0 \cup \mathcal{S}_1$ be a partition with $\mathcal{S}_0 \neq \emptyset$. Suppose that (a) $(T_{n,m})_{m\in\mathcal{S}_0} \xrightarrow{d} (Z_m)_{m\in\mathcal{S}_0}$ jointly, where $(Z_m)_{m\in\mathcal{S}_0}$ is a random vector with $P(|Z_m| < \infty) = 1$ for all $m \in \mathcal{S}_0$ and (b) for each $m \in \mathcal{S}_1$, either $P(T_{n,m} > C) \to 1$ for every constant $C > 0$, or $P(T_{n,m} < -C) \to 1$ for every constant $C > 0$. Then*

$$\min_{m=1,\ldots,M} |T_{n,m}| \xrightarrow{d} \min_{m\in\mathcal{S}_0} |Z_m|.$$

Proof of Lemma 1. Let G be a strictly increasing continuous distribution function on $\mathbb{R}$ extended to $\mathbb{R}_{[\pm\infty]} = \mathbb{R} \cup \{\pm\infty\}$ with $G(+\infty) = 1$ and $G(-\infty) = 0$. By condition (a) and the continuous mapping theorem, $(G(T_{n,m}))_{m\in\mathcal{S}_0} \xrightarrow{d} (G(Z_m))_{m\in\mathcal{S}_0}$, and by condition (b), either $G(T_{n,m}) \xrightarrow{p} 1$ or $G(T_{n,m}) \xrightarrow{p} 0$ for each $m \in \mathcal{S}_1$. Then, the joint convergence holds by Theorem 2.7(v) in van der Vaart (1998) and the following holds in $[0, 1]^M$,

$$G_n \equiv (G(T_{n,1}), \ldots, G(T_{n,M}))' \xrightarrow{d} G_\infty,$$

where G_∞ has components $G(Z_m)$ for $m \in \mathcal{S}_0$ and either 1 or 0 for $m \in \mathcal{S}_1$. Define $h : [0, 1]^M \to [0, +\infty]$ by $h(y) \equiv \min_{m=1,\ldots,M} |G^{-1}(y_m)|$, with the conventions $|G^{-1}(0)| = |G^{-1}(1)| = +\infty$, so that $\min_m |T_{n,m}| = h(G_n)$.

For any sequence $y^{(k)} \to y$, where any coordinate m with $y_m \in \{0, 1\}$, we have $|G^{-1}(y_m^{(k)})| \to +\infty$, so such coordinates do not affect the limit of the minimum, which is attained on $\mathcal{S}_0$. Therefore, the map h is continuous at $y \in [0, 1]^M \to \mathbb{R}^M_{[\pm\infty]}$. Thus, we have $h(G_n) \xrightarrow{d} h(G_\infty) = \min_{m\in\mathcal{S}_0} |Z_m|$ by the definition of $h(\cdot)$ and the continuous mapping theorem. □

The next lemma provides the properties of the limiting null distribution function that are used to establish consistency of the bootstrap critical value. It allows the correlation matrix of the limiting Gaussian vector to be singular, which occurs, for example, whenever the projection residuals $\widetilde{x}_{m,i}$ coincide across two correctly specified submodels.

Lemma 2. *Let $\mathcal{S}$ be a finite nonempty index set and let $(Z_m)_{m\in\mathcal{S}}$ be a mean zero Gaussian vector with $E[Z_m^2] = 1$ for all $m \in \mathcal{S}$. Then $F(x) \equiv P(\min_{m\in\mathcal{S}} |Z_m| \le x)$ satisfies the followings (i) F is continuous on $\mathbb{R}$ and $F(x) = 0$ for $x \le 0$; and (ii) F is strictly increasing on $[0, \infty)$. Consequently, F maps $[0, \infty)$ continuously and strictly increasingly onto $[0, 1)$, so for every $\alpha \in (0, 1)$ there exists a unique $c_{1-\alpha} \in (0, \infty)$ with $F(c_{1-\alpha}) = 1 - \alpha$, and the inverse quantile function $F^{-1}(\cdot)$ is continuous at $1 - \alpha$.*

Proof of Lemma 2. (i) is almost immediate since each Z_m is a nondegenerate standard normal random variable. For (ii), it suffices to show that the support of $W \equiv g((Z_m)_{m\in\mathcal{S}})$ contains $[0, \infty)$, where $g(z) \equiv \min_{m\in\mathcal{S}} |z_m|$: given $0 \le x_1 < x_2$, the midpoint $c = (x_1 + x_2)/2$ is then a support point of W and (x_1, x_2) is a neighborhood of c, so that $F(x_2) - F(x_1) \ge P(W \in (x_1, x_2)) > 0$. Let $\mathcal{Z}$ denote the support of $(Z_m)_{m\in\mathcal{S}}$, which is a linear subspace since the vector is mean zero Gaussian. Since $P((Z_m)_{m\in\mathcal{S}} \in \mathcal{Z}) = 1$ and $P(W > 0) = 1$, as $P(Z_m = 0) = 0$ for each of the finitely many $m \in \mathcal{S}$, the intersection of these two events is nonempty and yields a point $z^0 \in \mathcal{Z}$ with $g(z^0) > 0$. As $\mathcal{Z}$ is closed under scalar multiplication and g is positively homogeneous, $tz^0 \in \mathcal{Z}$ and $g(tz^0) = t\, g(z^0)$ for every $t > 0$, so g attains every value in $(0, \infty)$ on $\mathcal{Z}$; since g is continuous, each such value lies in the support of W, which therefore contains $(0, \infty)$ and hence, being closed, all of $[0, \infty)$. The final claim follows from (i) and (ii) together with $F(x) \to 1$ as $x \to \infty$. $\square$

Proof of Theorem 1

We first introduce notation used throughout the proof. Recall $q_i = (x_i, w_i', z_i')'$, and let $q_{m,i} \equiv (x_i, w_i', z_{m,i}')'$ for submodel m. Let $\widetilde{x}_{m,i}$ denote the residual from the population linear projection of x_i onto $(w_i', z_{m,i}')'$, and $\widehat{\widetilde{x}}_{m,i}$ its sample counterpart. Recall $Q_m = E[\widetilde{x}_{m,i}^2]$, $R_m = E[\widetilde{x}_{m,i} z_{-m,i}']$, and $\beta_m = Q_m^{-1} R_m \gamma_{-m}$ under H_0. Since $\widetilde{x}_{m,i}$ is a linear combination $c_m' q_{m,i}$ with coefficient 1 on x_i, Assumption 1(i) implies $Q_m = c_m' E[q_{m,i} q_{m,i}'] c_m \ge \underline{\lambda} > 0$ uniformly over $m = 1, \dots, M$. Let θ_m denote the population linear projection coefficients of y_i on $q_{m,i}$, and define the submodel projection error $u_{m,i} \equiv y_i - q_{m,i}' \theta_m$, which satisfies $E[q_{m,i} u_{m,i}] = 0$ by construction. Under $H_0 : \beta = 0$, write $y_i = w_i' \delta + z_{m,i}' \gamma_{(m)} + z_{-m,i}' \gamma_{-m} + e_i$, where $\gamma_{(m)}$ and γ_{-m} partition γ into the coefficients of the included and omitted auxiliary regressors. By the L^2 projection, $\mathrm{Proj}(y_i \mid q_{m,i}) = w_i' \delta + z_{m,i}' \gamma_{(m)} + \mathrm{Proj}(z_{-m,i}' \gamma_{-m} \mid q_{m,i})$ where we use $E[e_i q_{m,i}] = 0$ under Assumption 1(i). Subtracting yields the decomposition

$$u_{m,i} = e_i + a_{m,i}, \qquad a_{m,i} = z_{-m,i}' \gamma_{-m} - \mathrm{Proj}(z_{-m,i}' \gamma_{-m} \mid q_{m,i}),$$

where $a_{m,i}$ is a function of the regressors only and $u_{m,i} = e_i$ for every $m \in \mathcal{S}_0 = \{m : \gamma_{-m} = 0\}$. We write $\mathcal{S}_1 = \{1, \dots, M\} \setminus \mathcal{S}_0$ for the set of misspecified submodels, which satisfy $\beta_m \ne 0$ by Assumption 1(ii). Define $\Omega_{jl}^u \equiv E[\widetilde{x}_{j,i} \widetilde{x}_{l,i}\, u_{j,i} u_{l,i}]$, $\Omega_m^u \equiv \Omega_{mm}^u$, and $\sigma_{\beta,m}^2 \equiv Q_m^{-1} \Omega_m^u Q_m^{-1}$. For $j, l \in \mathcal{S}_0$, we have $\Omega_{jl}^u = \Omega_{jl} = E[\widetilde{x}_{j,i} \widetilde{x}_{l,i}\, e_i^2]$, as defined in the main text.

First of all, we observe that

$$\begin{aligned}
&\min_m |\widehat{\beta}_m / \widehat{\sigma}_m| > z_{1-\alpha/2} \\
&\Leftrightarrow |\widehat{\beta}_m / \widehat{\sigma}_m| > z_{1-\alpha/2} \quad \text{for all } m \\
&\Leftrightarrow 0 \notin [\widehat{\beta}_m - z_{1-\alpha/2} \widehat{\sigma}_m, \widehat{\beta}_m + z_{1-\alpha/2} \widehat{\sigma}_m] \quad \text{for all } m \\
&\Leftrightarrow 0 \notin \bigcup_m [\widehat{\beta}_m - z_{1-\alpha/2} \widehat{\sigma}_m, \widehat{\beta}_m + z_{1-\alpha/2} \widehat{\sigma}_m] \\
&\Longleftarrow 0 \notin [\min_m (\widehat{\beta}_m - z_{1-\alpha/2} \widehat{\sigma}_m), \max_m (\widehat{\beta}_m + z_{1-\alpha/2} \widehat{\sigma}_m)]
\end{aligned}$$

where the last line follows from $\bigcup_m[\widehat{\beta}_m - z_{1-\alpha/2}\widehat{\sigma}_m, \widehat{\beta}_m + z_{1-\alpha/2}\widehat{\sigma}_m] \subseteq [\min_m(\widehat{\beta}_m - z_{1-\alpha/2}\widehat{\sigma}_m), \max_m(\widehat{\beta}_m + z_{1-\alpha/2}\widehat{\sigma}_m)]$, but the converse does not hold in general.

We now derive the joint limiting behavior of the sample t-statistics. Under $H_0 : \beta = 0$ with the model (1) and Assumption 1(i),

$$\beta_m = E[\widetilde{x}_{m,i}^2]^{-1}E[\widetilde{x}_{m,i}y_i] = Q_m^{-1}R_m\gamma_{-m},$$

using $E[\widetilde{x}_{m,i}e_i] = 0$. By the Frisch-Waugh-Lovell theorem, $\widehat{\beta}_m = (\sum_i \widehat{\widetilde{x}}_{m,i}^2)^{-1}\sum_i \widehat{\widetilde{x}}_{m,i}\, y_i$, and standard arguments under Assumption 1(i) yield the following asymptotically linear representation for each $m = 1, \dots, M$,

$$\sqrt{n}(\widehat{\beta}_m - \beta_m) = Q_m^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{x}_{m,i}\, u_{m,i} + o_p(1),$$

where the summands are i.i.d. with mean zero and finite variance $\Omega_m^u < \infty$ by Assumption 1(i) and the Cauchy-Schwarz inequality. Under Assumption 1(i), $\Omega_m^u = E[\widetilde{x}_{m,i}^2 a_{m,i}^2] + E\big[\widetilde{x}_{m,i}^2\, E[e_i^2 \mid q_i]\big] \geq \underline{\sigma}^2 Q_m > 0$, where the cross term vanishes because $a_{m,i}$ is a function of q_i and $E[e_i \mid q_i] = 0$ by Assumption 1(i).

Then, by the Cramér-Wold device and the multivariate CLT, we have the following joint distribution;

$$\big(\sqrt{n}(\widehat{\beta}_m - \beta_m)\big)_{m\in\mathcal{S}_0} \xrightarrow{d} N(0, V), \qquad V_{jl} = Q_j^{-1}\Omega_{jl}^u Q_l^{-1}.$$

As we discussed above, for $m \in \mathcal{S}_0$, $\beta_m = 0$ and $u_{m,i} = e_i$, so the following holds jointly,

$$(t_m)_{m\in\mathcal{S}_0} = (\widehat{\beta}_m/\widehat{\sigma}_m)_{m\in\mathcal{S}_0} \xrightarrow{d} (Z_m)_{m\in\mathcal{S}_0},$$

where (Z_m) is the mean-zero Gaussian vector defined in Section 3 with $E[Z_jZ_l] = \Omega_{jl}/(\Omega_{jj}^{1/2}\Omega_{ll}^{1/2})$. Note that the heteroskedasticity-robust variance estimator is consistent $n\widehat{\sigma}_m^2 \xrightarrow{p} \sigma_{\beta,m}^2 = Q_m^{-1}\Omega_m^u Q_m^{-1}$ for all m by finite moment conditions in Assumption 1 and LLN.

For $m \in \mathcal{S}_1$, we decompose $t_m = \sqrt{n}(\widehat{\beta}_m - \beta_m)/\sqrt{n}\widehat{\sigma}_m + \sqrt{n}\beta_m/\sqrt{n}\widehat{\sigma}_m$. The first term is $O_p(1)$, while the second diverges to $\mathrm{sign}(\beta_m)\cdot(+\infty)$, since $\beta_m \neq 0$ by Assumption 1(ii) and $\sqrt{n}\,\widehat{\sigma}_m \xrightarrow{p} \sigma_{\beta,m} \in (0, \infty)$. Hence, for $m \in \mathcal{S}_1$, either $P(t_m > C) \to 1$ for every constant $C > 0$, or $P(t_m < -C) \to 1$ for every constant $C > 0$, according to the sign of β_m. Therefore, conditions (a) and (b) of Lemma 1 hold as shown above, and the following holds

$$\min_{m=1,\dots,M}|t_m| \xrightarrow{d} \min_{m\in\mathcal{S}_0}|Z_m|.$$

by Lemma 1. Then, it follows under H_0 that

$$\lim_{n\to\infty} P\big(\min_{m=1,\dots,M}|\widehat{\beta}_m/\widehat{\sigma}_m| > z_{1-\alpha/2}\big) = P\big(\min_{m\in\mathcal{S}_0}|Z_m| > z_{1-\alpha/2}\big) \leq P(|Z_{m^*}| > z_{1-\alpha/2}) = \alpha,$$

for any $m^* \in \mathcal{S}_0$, and this establishes (6). Together with the derivation above, we can also show the original EBA based on the extreme bounds interval is at least as conservative.

We now provide the validity of the wild bootstrap critical value. Under $H_0 : \beta = 0$ and Assumption 1(i), the true DGP is $y_i = w_i'\delta + z_i'\gamma + e_i$ with $E[e_i \mid q_i] = 0$, and the restricted OLS estimates satisfy $(\widetilde{\delta}, \widetilde{\gamma}) \xrightarrow{p} (\delta, \gamma)$, with restricted residuals $\widetilde{e}_i = e_i - w_i'(\widetilde{\delta} - \delta) - z_i'(\widetilde{\gamma} - \gamma)$. The wild bootstrap generates $y_i^* = w_i'\widetilde{\delta} + z_i'\widetilde{\gamma} + \widetilde{e}_i v_i$ with the regressors held fixed. By the Frisch-Waugh-Lovell theorem and the orthogonality of $\widehat{\widetilde{x}}_{m,i}$ to $(w_i', z_{m,i}')'$, we have $\widehat{\beta}_m^* = \check{\beta}_m + \Big(\sum_i \widehat{\widetilde{x}}_{m,i}^2\Big)^{-1} \sum_i \widehat{\widetilde{x}}_{m,i}\, \widetilde{e}_i v_i$,

$$\check{\beta}_m \equiv \Big(\frac{1}{n}\sum_i \widehat{\widetilde{x}}_{m,i}^2\Big)^{-1} \Big(\frac{1}{n}\sum_i \widehat{\widetilde{x}}_{m,i}\, z_{-m,i}'\Big)\widetilde{\gamma}_{-m} \xrightarrow{p} Q_m^{-1} R_m \gamma_{-m} = \beta_m.$$

So, the bootstrap reproduces the population omitted variable bias for all submodels: $\check{\beta}_m \xrightarrow{p} 0$ for $m \in \mathcal{S}_0$ and $\check{\beta}_m \xrightarrow{p} \beta_m \neq 0$ for $m \in \mathcal{S}_1$.

Conditional on the data, $\mathbb{Z}_{n,m}^* \equiv n^{-1/2}\sum_i \widehat{\widetilde{x}}_{m,i}\, \widetilde{e}_i v_i$ is a sum of independent mean-zero terms with conditional covariance $\frac{1}{n}\sum_i \widehat{\widetilde{x}}_{j,i}\widehat{\widetilde{x}}_{l,i}\, \widetilde{e}_i^2 \xrightarrow{p} E[\widetilde{x}_{j,i}\widetilde{x}_{l,i}\, e_i^2] = \Omega_{jl}, j, l \in \mathcal{S}_0$, by the LLN, and $E\|q_i\|^4 < \infty$, and $E[e_i^4] < \infty$ under Assumption 1. Conditional on the data, $(\mathbb{Z}_{n,m}^*)_{m\in\mathcal{S}_0} \xrightarrow{d^*} N(0, \Omega)$ with $\Omega \equiv (\Omega_{jl})_{j,l\in\mathcal{S}_0}$. For any $\lambda \in \mathbb{R}^{|\mathcal{S}_0|}$ and let $d_{n,i} \equiv \big(\sum_{m\in\mathcal{S}_0} \lambda_m \widehat{\widetilde{x}}_{m,i}\big)\widetilde{e}_i$, so that $\lambda'(\mathbb{Z}_{n,m}^*)_{m\in\mathcal{S}_0} = n^{-1/2}\sum_i d_{n,i} v_i$ has conditional mean zero and conditional variance $n^{-1}\sum_i d_{n,i}^2 \xrightarrow{p} \lambda'\Omega\lambda$. If $\lambda'\Omega\lambda = 0$, the Chebyshev inequality gives $n^{-1/2}\sum_i d_{n,i} v_i \xrightarrow{p^*} 0$. If $\lambda'\Omega\lambda > 0$, we apply the Lyapunov CLT for triangular arrays, where the Lyapunov condition can be verified with $\delta = 1$:

$$\sum_{i=1}^n E^*\big|n^{-1/2} d_{n,i} v_i\big|^3 = E|v_i|^3\, \frac{1}{n^{3/2}} \sum_{i=1}^n |d_{n,i}|^3 \leq E|v_i|^3 \Big(\frac{1}{n}\max_{i\leq n} d_{n,i}^2\Big)^{1/2}\Big(\frac{1}{n}\sum_{i=1}^n d_{n,i}^2\Big),$$

where $E|v_i|^3 < \infty$ by Assumption 1(iii) and the last factor is $O_p(1)$ by the conditional convergence of covariance shown above. We also have $\max_{i\leq n} d_{n,i}^2 \lesssim \max_{i\leq n}\|q_i\|^2 e_i^2 + o_p(1)\max_{i\leq n}\|q_i\|^4$ by using a form of $\widehat{\widetilde{x}}_{m,i}$ and the expansion of $\widetilde{e}_i$. Under Assumption 1(i), $E[\|q_i\|^2 e_i^2] < \infty$ by the Cauchy-Schwarz inequality and $E\|q_i\|^4 < \infty$, and thus $n^{-1}\max_{i\leq n} d_{n,i}^2 = o_p(1)$ and hence the Lyapunov condition holds. By the Cramér-Wold device, together with $n\widehat{\sigma}_m^{*2} \xrightarrow{p^*} Q_m^{-1}\Omega_{mm}Q_m^{-1}$ for $m \in \mathcal{S}_0$ (by the same arguments applied to the bootstrap residuals, using conditional Markov inequalities and $E[v_i^4] < \infty$), the bootstrap t-statistics jointly satisfy

$$(t_m^*)_{m\in\mathcal{S}_0} = (\widehat{\beta}_m^*/\widehat{\sigma}_m^*)_{m\in\mathcal{S}_0} \xrightarrow{d^*} (Z_m)_{m\in\mathcal{S}_0},$$

to the same limiting Gaussian vector (Z_m) as above. For $m \in \mathcal{S}_1$, $\widehat{\beta}_m^* = \check{\beta}_m + O_{p^*}(n^{-1/2})$ with $\check{\beta}_m \xrightarrow{p} \beta_m \neq 0$, while $\widehat{\sigma}_m^* = O_{p^*}(n^{-1/2})$. Hence, either $P^*(t_m^* > C) \xrightarrow{p} 1$ for every constant $C > 0$, or $P^*(t_m^* < -C) \xrightarrow{p} 1$ for every constant $C > 0$, according to the sign of β_m.

By Lemma 2 with $\mathcal{S} = \mathcal{S}_0$, we have that the CDF of limiting null distribution $F_0(x) \equiv P(\min_{m\in\mathcal{S}_0}|Z_m| \leq x)$ is continuous on $\mathbb{R}$ and strictly increasing on $[0, \infty)$, and that for every $\alpha \in (0, 1/2)$ there exists a unique $c_{1-\alpha} \in (0, \infty)$ with $F_0(c_{1-\alpha}) = 1 - \alpha$, at which the inverse quantile function $F_0^{-1}(\cdot)$ is continuous.

Every subsequence $\{u_n\}$ of $\{n\}$ contains a further subsequence along which the finitely many convergence in probability results established above (consistency of the restricted estimates and the sample projection coefficients, convergence of the conditional covariances, $\check{\beta}_m \xrightarrow{p} \beta_m$ for all m, and convergence of the bootstrap variance estimators) hold almost surely. Fix such a further subsequence and a sample path in the corresponding probability one event. Then, conditional on the data and conditional on this subsequence, conditions (a) and (b) of Lemma 1 hold for the bootstrap t-statistics: (a) $(t_m^*)_{m\in\mathcal{S}_0} \xrightarrow{d} (Z_m)_{m\in\mathcal{S}_0}$, and (b) $P^*(t_m^* > C) \to 1$ or $P^*(t_m^* < -C) \to 1$ for each $m \in \mathcal{S}_1$ and every $C > 0$. Therefore, we have

$$\widehat{F}_n^*(x) \equiv P^*\Big(\min_{m=1,\ldots,M} |\widehat{\beta}_m^*/\widehat{\sigma}_m^*| \le x\Big) \to F_0(x) \quad \text{for all } x \in \mathbb{R}$$

conditional on the subsequence. Since every subsequence contains such a further subsequence, the subsequence characterization of convergence in probability gives $\widehat{F}_n^*(x) \xrightarrow{p} F_0(x)$ for each fixed $x \in \mathbb{R}$, which in turn implies the (unconditional) convergence of the bootstrap statistics $\min_m |t_m^*| \xrightarrow{d} \min_{m\in\mathcal{S}_0} |Z_m|$. Note that $\widehat{c}_{1-\alpha} = \widehat{F}_n^{*-1}(1-\alpha) = \inf\{x : \widehat{F}_n^*(x) \ge 1-\alpha\}$. Since F_0^{-1} is continuous at $1-\alpha$, by Lemma 21.2 of van der Vaart (1998), we have that the conditional bootstrap quantile satisfies $\widehat{c}_{1-\alpha} \xrightarrow{p} c_{1-\alpha}$. Combining this with $\min_m |t_m| \xrightarrow{d} \min_{m\in\mathcal{S}_0} |Z_m|$, we have under H_0

$$\lim_{n\to\infty} P\big(\min_{m=1,\ldots,M} |\widehat{\beta}_m/\widehat{\sigma}_m| > \widehat{c}_{1-\alpha}\big) = P\big(\min_{m\in\mathcal{S}_0} |Z_m| > c_{1-\alpha}\big) = \alpha.$$

This completes the proof. *Q.E.D.*

B Variable Descriptions

Here, we describe the main variables we use in the cross-country growth regressions in Section 5. See also the Data Appendix in Levine and Renelt (1992) for more detailed descriptions.

1. Dependent Variables

    - GDP: average yearly growth rate of real per capita gross domestic product.
    - INV: average investment share as a GDP portion (also a core variable for GDP)

2. Core Variables

    - RGDP60: the Real GDP per capita in 1960
    - GPO: the average Growth rate of population (demographics)
    - SEC: secondary school enrollment rate in 1960 (in 1970)

3. The Fiscal Policy variables
    - GOV: the average Government consumption expenditures (as GDP share),
    - GOVX: Government consumption share (of GDP) net of Defense and Educational Expenses
    - TEX: Ratio of total government Expenditure share to GDP
    - DEF: Ratio of central-government deficit to GDP
4. Trade Variables
    - EXP: the average Exports (as GDP share),
    - IMP: the average Imports (as GDP share),
    - LEAM1: Leamer's (1988) openness index on factor-adjusted trade, accounting for the difference between actual and predicted level of trade,
    - LEAM2: Leamer's (1988) trade-distortion/intervention measure, based on the classical Heckscher-Ohlin deviations between actual and predicted pattern of trade,
    - BMP: the average Black-Market exchange-rate Premium
    - RERDB: Dollar's (1992) Real Exchange-Rate Distortion between domestic and international prices.
5. Monetary and Political Variables
    - PI: the average Inflation of GDP deflator
    - GDC: the average Growth rate of Domestic Credit
    - STDI: the Standard deviation of the rate of inflation
    - STDD: the Standard deviation of Domestic Credit Growth
    - REVC: counter of revolutions and coups per year, see Barro (1991)