

# Confidence Intervals for Bias and Size Distortion in IV and Local Projections–IV Models

Gergely Ganics<sup>1</sup>, Atsushi Inoue<sup>2</sup> and Barbara Rossi<sup>3</sup>

September 29, 2018

## Abstract

In this paper we propose methods to construct confidence intervals for the bias of the two-stage least squares estimator, and the size distortion of the associated Wald test in instrumental variable models. Importantly our framework covers the local projections – instrumental variable model as well. Unlike tests for weak instruments, whose distributions are non-standard and depend on nuisance parameters that cannot be estimated consistently, the confidence intervals for the strength of identification are straightforward and computationally easy to calculate, as they are obtained from inverting a chi-squared distribution. Furthermore, they provide more information to researchers on instrument strength than the binary decision offered by tests. Monte Carlo simulations show that the confidence intervals have good small sample coverage. We illustrate the usefulness of the proposed methods to measure the strength of identification in two empirical situations: the estimation of the intertemporal elasticity of substitution in a linearized Euler equation, and government spending multipliers.

**Keywords:** Instrumental Variables, Weak Instruments, Weak Identification, Concentration Parameter, Local Projections.

**J.E.L. Codes:** C22, C52, C53.

---

<sup>1</sup>ADG Economics and Research, Bank of Spain, C/Alcalá 48, Madrid 28014, Spain. E-mail: gergely.ganics@bde.es.

<sup>2</sup>Department of Economics, Vanderbilt University, VU Station B, Box #351819, 2301 Vanderbilt Place, Nashville, TN 37235, USA. E-mail: atsushi.inoue@vanderbilt.edu.

<sup>3</sup>Corresponding author: ICREA-Univ. Pompeu Fabra, CREI and Barcelona GSE, c/Ramon Trias Fargas, 25-27, Mercè Rodoreda bldg., 08005 Barcelona, Spain. E-mail: barbara.rossi@upf.edu.

Acknowledgments: Gergely Ganics gratefully acknowledges financial support from the Spanish Ministerio de Economía y Competitividad under FPI grant BES-2013-065352. Barbara Rossi gratefully acknowledges financial support from the Fundación BBVA scientific research grant (PR16\_DAT\_0043) on Analysis of Big Data in Economics and Empirical Applications, the Spanish Ministry of Economy and Competitiveness, Grant ECO2015-68136-P and FEDER, UE and the Cerca Programme/Generalitat de Catalunya; the Barcelona GSE acknowledges financial support from the Spanish ministry of the Economy and Competitiveness through the Severo Ochoa Programme for Centers of Excellence in R&D (SEV-2015-0563). We thank José Luis Montiel Olea and Carolin Pflueger for sharing their code, Majid Al-Sadoon, as well as participants at the 2017 UPF Workshop on Forecasting and Time Series Econometrics, the 2017 Barcelona GSE Summer Forum in Time Series and the Bank of Spain for comments. The views expressed herein are those of the authors and should not be attributed to the Bank of Spain or the Eurosystem.

# 1 INTRODUCTION

In this paper, we propose a novel methodology to construct confidence intervals for the strength of identification, and in particular the bias and size distortion in linear instrumental variables (IV) models. Measuring the strength of identification is an extremely important issue in practice. It is well-known that the presence of weak instruments invalidates standard inference (Stock, Wright and Yogo, 2002), leading to inconsistent point estimates, incorrectly sized tests and invalid confidence intervals. A conventional and widely-used approach to detect weak instruments in practice is using the first-stage  $F$ -statistic, which is the  $F$ -statistic on the strength of the instrument identification. The statistic (or its generalization, in the case of multiple endogenous regressors) was proposed by Staiger and Stock (1997), Stock, Wright and Yogo (2002), Stock and Yogo (2005) and Montiel Olea and Pflueger (2013) as an approach to evaluate the severity of the weak instrument problem in specific empirical applications. A large enough value of the first-stage  $F$ -statistic (judged according to appropriately derived critical values) increases researchers' confidence that the instruments are strong and, thus, that standard inference on the structural parameters of interest is valid. Our complementary approach is instead based on constructing a confidence interval for the strength of identification in terms of quantities of primary interest: bias and size distortion in the homoskedastic IV model, and bias in the heteroskedastic/autocorrelated IV model with one endogenous variable as well as in the local projections–IV (LP-IV) framework.

From a practical point of view, as Stock, Wright and Yogo (2002, p. 518) point out, "Finding exogenous instruments is hard work, and the features that make an instrument plausibly exogenous, such as occurring sufficiently far in the past to satisfy a first-order condition or the as-if random coincidence that lies behind a quasi-experiment, can also work to make the instrument weak." Once a researcher has gone through the tedious job of finding exogenous instruments, he or she can rely on our method to *quantify* potential issues caused by the specific instruments' strength, without having to discard the instruments altogether.

From a methodological perspective, confidence intervals and other statistics reflecting sampling uncertainty provide additional information relative to  $p$ -values, recently urged by the American Statistical Association (Wasserstein and Lazar, 2016) and also demanded by the economics community (e.g. the American Economic Review's Submission Guidelines state "In tables, please report standard errors in parentheses but do not use \*s to report significance levels.")

In our frameworks, the strength of identification as well as the bias of the two-stage least squares (TSLS) estimator and the size distortion of the associated Wald test depend on two types of parameters: coefficients which cannot be consistently

estimated and covariances which are consistently estimable. Our proposed procedure works as follows. In the first step, we construct asymptotically valid  $(1 - \alpha)$  level confidence sets for the former set of parameters. The second step depends on the model. In the homoskedastic IV model, we form the confidence intervals for the parameter summarizing the strength of identification by using the aforementioned confidence sets and plugging the consistent covariance estimates in the appropriate expression for the strength of identification. To construct confidence intervals for the bias and the size distortions, we exploit the mapping from the parameter summarizing the strength of identification to bias or size distortion via the projection method – see e.g. Dufour (1997) for an early application of the projection method with weak instruments. In particular, in the case of one endogenous regressor in the homoskedastic IV model, we can construct our confidence intervals for the strength of identification based on the non-central chi-squared distribution, resulting in tight confidence intervals whose coverage rates are very close to their nominal level. In the heteroskedastic/autocorrelated IV model, we utilize the confidence sets and consistent estimates from the first step to obtain confidence intervals for the Nagar (1959) bias directly through the projection method. We note that in general, the projection method leads to conservative confidence intervals.

The methodology that we propose has several attractive properties. First, it provides guidance to applied researchers on *quantifying* the strength of instruments as well as bias and size distortion in their empirical analyses, and thus protects them against weak instruments. A second advantage is that the confidence intervals for the strength of identification are straightforward and computationally easy to calculate, as they are obtained from inverting asymptotic chi-squared distributions. The simplicity of our confidence intervals distinguishes our methodology from weak instrument tests, whose distributions are typically asymptotically non-pivotal and depend on nuisance parameters that cannot be estimated consistently.

A third advantage of our methodology is that it can be applied in the presence of heteroskedasticity and serial correlation when there is one endogenous regressor. Our framework is also general enough to be applied to LP-IV models (Jordà, 2005). Since the construction of confidence intervals for the strength of identification is based on inverting an asymptotic chi-squared distribution, the methodology can be easily applied even if the disturbances are heteroskedastic and/or serially correlated, in which case one will simply use a Heteroskedasticity and Autocorrelation Consistent (HAC) estimator.

Monte Carlo simulations demonstrate that our methods have good coverage.

We illustrate the usefulness of our methodology in two empirical applications. In the first one, we estimate the intertemporal elasticity of substitution in linearized Euler equations in a heteroskedastic/autocorrelated IV model, following Yogo (2004)

and Montiel Olea and Pflueger (2013). Our confidence intervals confirm that weak identification is indeed a serious problem, preventing reliable estimation of the intertemporal elasticity of substitution. In the second empirical application, we analyze the identification of a local projections-IV model to estimate government spending multipliers, following Ramey and Zubairy (2018).

Our paper is related to the literature on testing the strength of instruments in linear IV models, in particular Staiger and Stock (1997), Stock, Wright and Yogo (2002), and Stock and Yogo (2005), who discuss the use of a first-stage  $F$ -statistic to test whether instruments are weak, and Montiel Olea and Pflueger (2013), who provide the limiting distribution of an appropriate first-stage  $F$ -statistic under heteroskedasticity and serial correlation when there is only one included endogenous variable. We also make a methodological contribution by constructing confidence intervals for the bias of the local projections-IV estimator proposed by Jordà (2005).

An alternative approach would be to construct confidence intervals robust to weak identification for the structural parameters, a solution that becomes computationally infeasible in large dimensional settings and is only available in special cases. Tests for weak instruments can be less computationally challenging and are widely used in practice for their simplicity. Thus, the confidence intervals for the bias and size distortion that we propose are a practically convenient complementary approach to robust inference methodologies.

The paper is organized as follows. [Section 2](#) provides the intuition behind our method using a simple example. [Section 3](#) describes the econometric frameworks we consider and our proposed confidence intervals. [Section 4](#) provides Monte Carlo simulation results. [Section 5](#) presents empirical results, and [Section 6](#) concludes.

## 2 AN ILLUSTRATIVE EXAMPLE

This section illustrates the intuition behind our results in the context of a simple example. Consider the following baseline IV model:

$$y = Y\beta + u, \tag{1}$$

$$Y = Z\Pi + V, \tag{2}$$

where  $y$  is a  $(T \times 1)$  vector,  $T$  is the sample size,  $Y$  is a  $(T \times 1)$  vector of included endogenous variables and  $Z$  is a  $(T \times 1)$  vector of instruments (excluded exogenous variables);  $u$  and  $V$  are  $(T \times 1)$  vectors of independent, mean-zero disturbances with variance  $\sigma_{uu}$  and  $\sigma_{VV}$ , where, for simplicity,  $\sigma_{uu}$  and  $\sigma_{VV}$  are known, along with  $E(z_t^2)$ . The structural equation is [eq. \(1\)](#), with the structural coefficient of interest  $\beta$ . Information on the strength of the instrument is carried by the parameter  $\Pi$  in [eq. \(2\)](#).

## 2.1 Confidence Intervals for the Strength of Identification

In the case of one endogenous variable, the standard tests for the strength of instruments rely on Stock and Yogo (2005) and Stock, Wright and Yogo (2002), who recommend using a first-stage  $F$ -statistic. This statistic is formally constructed as a test of the null hypothesis that the instrument is not correlated with the endogenous variable ( $\Pi = 0$ ) against the alternative that  $\Pi \neq 0$ . The aforementioned papers derive the distribution of the first-stage  $F$ -statistic under the assumption that instruments are weak, that is  $\Pi = C/\sqrt{T}$ , where  $C$  is a constant, for testing the null hypothesis that the instrument strength is less than or equal to a threshold against the alternative that it exceeds the threshold. In this approach, the asymptotic distribution of the test statistic is asymptotically non-pivotal, as it depends on a nuisance parameter ( $C$ ) that cannot be consistently estimated, and this parameter plays a central role in determining the bias of the TSLS estimator and the size distortion of its associated Wald test. Therefore test statistic's critical values are different from standard values based on the chi-squared distribution, thus making inference difficult.

Let  $\hat{\Pi}_T = (Z'Z)^{-1} (Z'Y)$  denote the Ordinary Least Squares (OLS) estimator of  $\Pi$  in eq. (2). The reason why the first-stage  $F$ -statistic,  $\mathcal{F}_0$ , is asymptotically non-pivotal is because, under the assumptions in Stock and Yogo (2005) and Staiger and Stock (1997):

$$\mathcal{F}_0 \equiv \frac{(\hat{\Pi}_T - 0)^2}{\sigma_{VV} (Z'Z)^{-1}} = \frac{\left[ \sqrt{T} (\hat{\Pi}_T - 0) \right]^2}{\sigma_{VV} \left( \frac{Z'Z}{T} \right)^{-1}} = Y'Z (Z'Z)^{-1} Z'Y \frac{1}{\sigma_{VV}}, \quad (3)$$

and

$$\begin{aligned} \sqrt{T} (\hat{\Pi}_T - 0) &= \sqrt{T} (Z'Z)^{-1} (Z'Y) = \left( \frac{Z'Z}{T} \right)^{-1} \left( \frac{Z'Y}{\sqrt{T}} \right) \\ &= \left( \frac{Z'Z}{T} \right)^{-1} \left( \frac{Z'ZC}{T} \right) + \left( \frac{Z'Z}{T} \right)^{-1} \left( \frac{Z'V}{\sqrt{T}} \right) \\ &\xrightarrow{d} C + \nu, \end{aligned} \quad (4)$$

where  $\nu = E(Z_t^2)^{-1} \Psi_{ZV}$ ,  $\Psi_{ZV}$  is a random variable whose distribution is  $\mathcal{N}(0, E(Z_t^2) \sigma_{VV})$ , and  $\xrightarrow{d}$  denotes convergence in distribution. Thus, since the limiting distribution in eq. (4) depends on  $C$ , the distribution of the first-stage  $F$ -statistic in eq. (3) depends on  $C$ . This argument can be extended to the case of multiple endogenous regressors and instruments.

In our case, we focus on constructing a confidence interval for  $C$ . Note that the

dependence of the limiting distribution on  $C$  disappears when considering:

$$\sqrt{T} (\hat{\Pi}_T - \Pi) = \sqrt{T} \left( (Z'Z)^{-1} (Z'Y) - \Pi \right) \quad (5)$$

$$= \left( \frac{Z'Z}{T} \right)^{-1} \left( \frac{Z'Z}{T} C \right) + \left( \frac{Z'Z}{T} \right)^{-1} \left( \frac{Z'V}{\sqrt{T}} \right) - C \xrightarrow{d} v. \quad (6)$$

This result implies that

$$\mathcal{F}_\Pi \equiv \frac{(\hat{\Pi}_T - \Pi)^2}{\sigma_{VV} (Z'Z)^{-1}} \xrightarrow{d} \chi_1^2, \quad (7)$$

where  $\chi_1^2$  denotes a chi-squared distribution with one degree of freedom. Thus, one conveniently obtains a confidence interval for  $C$  by inverting a standard  $\chi_1^2$  distribution.

It might be surprising that the confidence intervals that we propose can be obtained by inverting limiting standard chi-squared distributions while the test statistics are not. The intuition is that the first-stage  $F$ -statistics is based on the difference between the estimate of the strength of identification and zero (the value that corresponds to no identification); hence, the difference between the two contains information on the true strength of identification and how close to zero that is, which cannot be consistently estimated. Thus, deriving the limiting distribution of the first-stage  $F$ -statistic in the weak instrument case results in a limiting distribution that is non-pivotal and depends on a parameter that cannot be estimated consistently. Confidence intervals, instead, are based on the difference between the estimate and the true strength of identification, rather than its value under the null hypothesis, and the limiting distribution of such difference does not depend on how close to zero the strength of identification is. Interestingly, this rather peculiar feature of the weak instrument problem cannot be applied to other non-standard situations resulting from the fact that the parameter is local to the null hypothesis, such as confidence intervals for highly persistent (local-to-unity) autoregressive processes. The reason is that, in the local-to-unity framework, the difference between the estimated largest root and its true value is a function of the Ornstein-Uhlenbeck process that approximates the true autoregressive process itself. Since the Ornstein-Uhlenbeck process is a function of the local-to-unity parameter, the limiting distribution remains a function of the latter. In our weak instrument case, instead, the local-to-zero parameter does not affect the limiting distribution of the variables themselves.

## 2.2 Confidence Intervals for Bias and Size Distortion

In this paper, we show how to construct confidence intervals for functions of  $C$  which measure the strength of the instrument, such as the concentration parameter, bias and size distortion. In this subsection, we focus on the size distortion (similar results apply for the bias when it exists, i.e. in overidentified models). It is well-known (e.g. Stock and Yogo, 2005) that, in the example considered in this section, the size distortion of a Wald test on the TOLS estimator of  $\beta$  is a function of the concentration parameter:

$$\mu_1^2 = C^2 E(z_i^2) / \sigma_{VV}. \quad (8)$$

Let us define  $s(\mu_1^2)$  to be the size distortion, where the notation emphasizes that it is a function of the concentration parameter. [Figure 1](#) shows the size distortion as a function of  $\mu_1^2$  (the nominal level of the Wald test is 5%).

Note that once one has a confidence interval for  $C$ ,  $CI_C$ , one directly obtains a confidence interval for  $\mu_1^2$ ,  $CI_{\mu_1^2}$ , as follows:

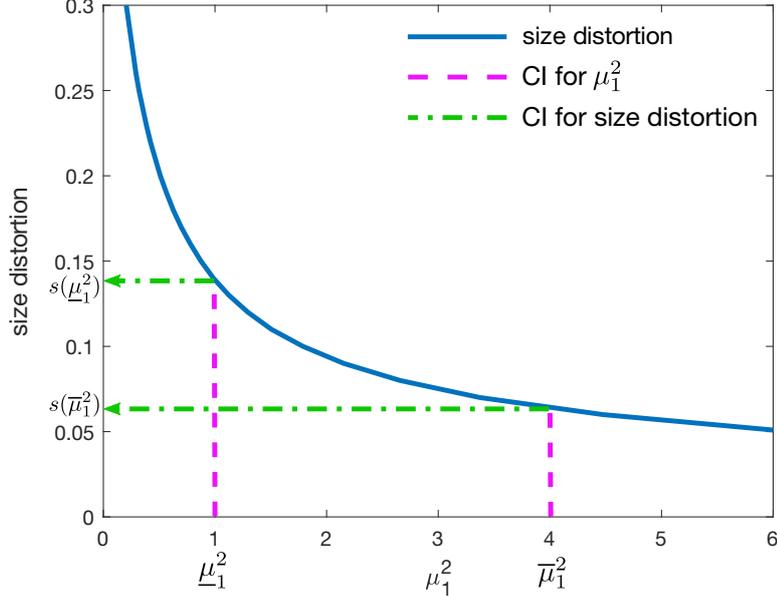
$$CI_{\mu_1^2} = \left\{ \tilde{\mu}_1^2 = \tilde{C}^2 E(z_i^2) / \sigma_{VV} \text{ such that } \tilde{C} \in CI_C \right\}. \quad (9)$$

One can then construct a confidence interval for  $s(\mu_1^2)$  by the projection method. Suppose  $CI_C = [\underline{C}, \overline{C}]$  is the confidence interval for  $C$  obtained by inverting the  $\chi_1^2$  distribution, where for simplicity  $\underline{C} > 0$ . Then  $CI_{\mu_1^2} = \left[ \underline{\mu}_1^2 = \underline{C}^2 E(z_i^2) / \sigma_{VV}, \overline{\mu}_1^2 = \overline{C}^2 E(z_i^2) / \sigma_{VV} \right]$ . Suppose that  $CI_{\mu_1^2} = [1, 4]$  is the confidence interval for  $\mu_1^2$ . Then the confidence interval for the size distortion,  $\left[ s\left(\underline{\mu}_1^2\right), s\left(\overline{\mu}_1^2\right) \right]$ , obtains as sketched in [Figure 1](#), and equals  $[0.06, 0.14]$ .

## 3 ECONOMETRIC FRAMEWORKS

In this section, we describe the three econometric frameworks we consider, and the corresponding confidence intervals that we propose. Throughout the paper,  $T$  denotes the sample size,  $\xrightarrow{p}$  and  $\xrightarrow{d}$  stand for convergence in probability and in distribution, respectively. The Euclidean norm of a vector  $a$  is denoted by  $\|a\|$ ,  $\text{tr}(\cdot)$  is the trace operator,  $\text{vec}(\cdot)$  is the vectorization operator, and  $\otimes$  is the Kronecker product. The abbreviation *iid* stands for independent and identically distributed,  $\mathcal{N}(\psi, \Xi)$  denotes the normal distribution with mean vector  $\psi$  and covariance matrix  $\Xi$ , and  $\chi_k^2$  denotes a chi-squared distribution with  $k$  degrees of freedom. For any  $(T \times K)$  matrix  $A$ ,  $P_A \equiv A(A'A)^{-1}A'$ , and  $M_A = I_K - P_A$ , where  $I_K$  is the  $(K \times K)$  identity matrix. We adopt the convention that for a symmetric positive definite matrix  $B$ ,  $B = B^{1/2}B^{1/2}$  and  $B^{-1} = B^{-1/2}B^{-1/2}$ , where  $B^{1/2}$  and  $B^{-1/2}$  are the unique principal square roots.

Figure 1: Construction of confidence interval for size distortion



Note: The figure plots size distortion as a function of  $\mu_1^2$  (solid line). The confidence interval for  $\mu_1^2$  is marked on the horizontal axis (vertical dashed lines), and the corresponding confidence interval for the size distortion is marked on the vertical axis (horizontal dash-dotted lines with arrows).

### 3.1 The Linear Homoskedastic IV Model

Consider the model of Staiger and Stock (1997) and Stock and Yogo (2005) (henceforth SSY), whose notation we follow:

$$y = Y\beta + X\gamma + u, \quad (10)$$

$$Y = Z\Pi + X\Phi + V, \quad (11)$$

where  $y$  is a  $(T \times 1)$  vector and  $Y$  is a  $(T \times n)$  matrix of included endogenous variables.  $X$  is a  $(T \times K_1)$  matrix of included exogenous variables (including a column of ones if there is a constant in eq. (10)) and  $Z$  is a  $(T \times K_2)$  matrix of excluded exogenous variables (instruments).  $\beta$  is an  $(n \times 1)$ , while  $\gamma$  is a  $(K_1 \times 1)$  vector of coefficients.  $\Pi$  is a matrix of coefficients of dimension  $(K_2 \times n)$ , and  $\Phi$  is a  $(K_1 \times n)$  matrix of coefficients. Furthermore,  $u$  is a  $(T \times 1)$  vector of errors, and  $V$  is a  $(T \times n)$  matrix of errors. Equation (10) is the structural equation of interest to the researcher and eq. (11) is the first stage equation relating the matrix of endogenous regressor(s)  $Y$  to the matrix of instrument(s)  $Z$ .

We define  $X_t = (X_{1t}, \dots, X_{K_1t})'$ ,  $Z_t = (Z_{1t}, \dots, Z_{K_2t})'$ ,  $V_t = (V_{1t}, \dots, V_{nt})'$ ,  $\underline{Z}_t = (X_t', Z_t')'$  as the vectors of the  $t$ -th observations of the respective variables,  $t = 1, \dots, T$ ,

and  $\underline{Z} = [X \ Z]$ . The population second moment matrices  $\Sigma$  and  $Q$  are as follows:

$$\Sigma = E \left[ \begin{pmatrix} u_t \\ V_t \end{pmatrix} \begin{pmatrix} u_t & V_t' \end{pmatrix} \right] = \begin{bmatrix} \sigma_{uu} & \Sigma_{uV} \\ \Sigma_{Vu} & \Sigma_{VV} \end{bmatrix}, \quad (12)$$

$$Q = E (\underline{Z}_t \underline{Z}_t') = \begin{bmatrix} Q_{XX} & Q_{XZ} \\ Q_{ZX} & Q_{ZZ} \end{bmatrix}. \quad (13)$$

In this section we make the same assumptions as SSY.

**Assumption  $L_\Pi$ :**  $\Pi = \Pi_T = C / \sqrt{T}$  where  $C$  is a fixed  $K_2 \times n$  matrix.

**Assumption M:** The following limits hold jointly for fixed  $K_2$  as  $T \rightarrow \infty$ :

(a)  $(T^{-1}u'u, T^{-1}V'u, T^{-1}V'V) \xrightarrow{p} (\sigma_{uu}, \Sigma_{Vu}, \Sigma_{VV});$

(b)  $T^{-1} \underline{Z}' \underline{Z} \xrightarrow{p} Q$ , where  $Q$  is positive definite;

(c)  $(T^{-1/2}X'u, T^{-1/2}Z'u, T^{-1/2}X'V, T^{-1/2}Z'V) \xrightarrow{d} (\Psi_{Xu}, \Psi_{Zu}, \Psi_{XV}, \Psi_{ZV})$ , where  $\Psi \equiv [\Psi'_{Xu}, \Psi'_{Zu}, \text{vec}(\Psi_{XV})', \text{vec}(\Psi_{ZV})']' \sim \mathcal{N}(0, \Sigma \otimes Q)$ , where  $\Sigma$  is positive definite.

Assumption  $L_\Pi$  models  $\Pi$  as local to zero, formalizing the weak instrument case, while Assumption M ensures that the appropriately scaled moments of the errors and the variables obey a Weak Law of Large Numbers and a Central Limit Theorem. Part (c) of Assumption M corresponds most naturally to serially uncorrelated and conditionally homoskedastic errors, which may be restrictive in certain empirical applications. This assumption will be substantially relaxed in [Section 3.2](#).

In order to develop our asymptotic theory, it is convenient to project out the exogenous regressors,  $X$ . That is, let  $Y^\perp \equiv M_X Y$ ,  $Z^\perp \equiv M_X Z$ , and  $V^\perp \equiv M_X V$ . Moreover, let  $V_t^\perp$  be the transpose of the  $t$ -th row of  $V^\perp$ , and similarly for  $Z_t^\perp$ . Note that, by the exogeneity of  $X$ ,  $E(X_t V_t') = 0$ , thus  $\Sigma_{V^\perp V^\perp} \equiv E(V_t^\perp V_t^{\perp'}) = \Sigma_{VV}$ . Using Assumption M, it can be shown that  $\hat{\Sigma}_{VV} \equiv Y^{\perp'} M_{Z^\perp} Y^\perp / (T - K_1 - K_2) \xrightarrow{p} \Sigma_{VV}$ . Using this notation, we can rewrite [eq. \(11\)](#) as:

$$Y^\perp = Z^\perp \Pi + V^\perp. \quad (14)$$

Furthermore, let us define  $\Omega \equiv Q_{ZZ} - Q_{ZX} Q_{XX}^{-1} Q_{XZ} = Q_{Z^\perp Z^\perp}$ , where  $Q_{Z^\perp Z^\perp} \equiv E(Z_t^\perp Z_t^{\perp'})$ , and  $\hat{\Omega} \equiv Z^{\perp'} Z^\perp / T$ . Moreover, let  $\hat{\Pi}_T \equiv (Z^{\perp'} Z^\perp)^{-1} Z^{\perp'} Y^\perp$  denote OLS estimator of  $\Pi$  in [eq. \(14\)](#). Note that Assumption M implies  $\hat{\Omega} \xrightarrow{p} \Omega$ .

The concentration matrix (parameter) plays an important role in the construction of the confidence intervals for the strength of identification. The concentration matrix  $\Lambda$  is given by

$$\Lambda \equiv \frac{1}{K_2} \Sigma_{VV}^{-1/2'} C' \Omega C \Sigma_{VV}^{-1/2} = \frac{1}{K_2} \lambda' \lambda, \quad (15)$$

where  $\lambda = \Omega^{1/2} C \Sigma_{VV}^{-1/2}$ . In the case of  $n = 1$  endogenous regressor,  $\Sigma_{VV}$  is a scalar  $\sigma_{VV}$  (whose consistent estimator is the same,  $\hat{\sigma}_{VV} = \hat{\Sigma}_{VV}$ ), and the concentration

matrix simplifies to the scalar concentration parameter:

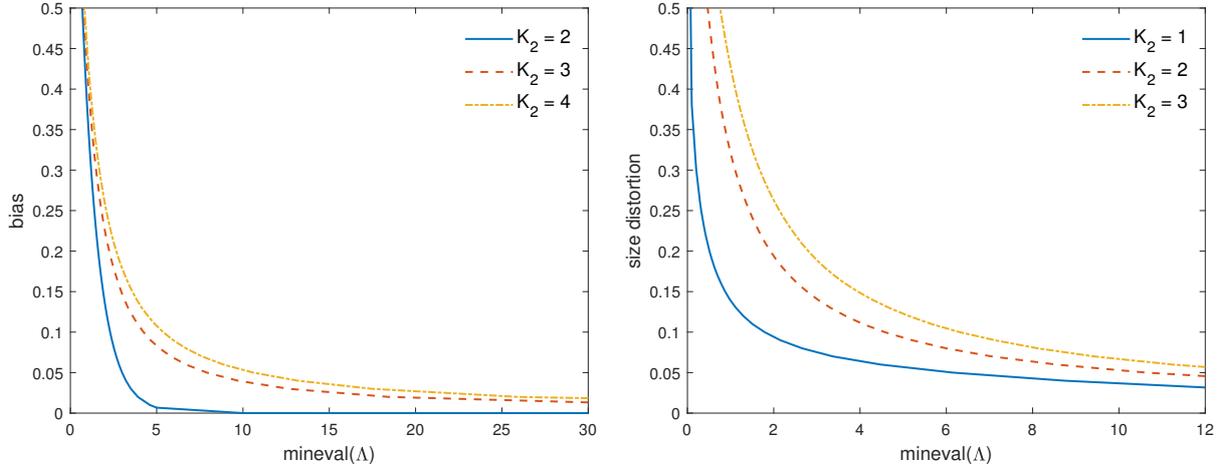
$$\mu_{K_2}^2 \equiv \frac{1}{K_2} C' \Omega C / \sigma_{VV}. \quad (16)$$

As Stock and Yogo (2005) demonstrated, the (i) worst-case asymptotic bias relative to the OLS estimator or (ii) worst-case asymptotic size distortion of the Wald-test on  $\beta$  – where the worst-case corresponds to the maximum of these quantities over all possible degrees of simultaneity between the error terms in eqs. (10) and (11) – of several  $k$ -class instrumental variables estimators, including the TSLS estimator, are functions (given  $n$  and  $K_2$ ) of the minimum eigenvalue of the concentration matrix, denoted by  $\text{mineval}(\Lambda)$ , or in the special case when  $n = 1$ , of the concentration parameter. In what follows, for simplicity we refer to (i) and (ii) as bias and size distortion, respectively. Furthermore, due to the popularity of the TSLS estimator, we will focus on it. In order to better understand the important role played by  $\Lambda$ , note that the matrix analog of the first-stage  $F$ -statistic testing the null hypothesis  $\Pi = 0$  is  $G_T = \frac{1}{K_2} \widehat{\Sigma}_{VV}^{-\frac{1}{2}} Y^\perp{}' P_{Z^\perp} Y^\perp \widehat{\Sigma}_{VV}^{-\frac{1}{2}}$ . The Cragg and Donald (1993) and Stock and Yogo (2005) test statistic,  $g_{\min}$ , is the minimum eigenvalue of  $G_T$ :  $g_{\min} = \text{mineval}(G_T)$ . As Stock and Yogo (2005) demonstrate,  $G_T \xrightarrow{d} \nu_1 / K_2$  and  $g_{\min} \xrightarrow{d} \text{mineval}(\nu_1 / K_2)$ , where  $\nu_1$  has a non-central Wishart distribution with non-centrality matrix  $\lambda' \lambda = K_2 \Lambda$ .

Let  $b(\text{mineval}(\Lambda); n, K_2)$  and  $s(\text{mineval}(\Lambda); n, K_2)$  denote the bias and the size distortion of the TSLS estimator, respectively, as the function of  $\text{mineval}(\Lambda)$  when the number of endogenous regressors and instruments are  $n$  and  $K_2$ , respectively, which we assume to be fixed. For general  $n$ , no closed-form expression is known for the functions  $b$  and  $s$ , although their values can be simulated following the algorithm given by Stock and Yogo (2005), suggesting they are continuous and decreasing. However, recently Skeels and Windmeijer (2016) obtained an expression for the bias function  $b$  for the case of  $n = 1$  endogenous variable. Figure 2 shows the simulated functions  $b$  and  $s$  for  $n = 1$  endogenous regressor and various numbers of instruments  $K_2$ . Section D of the Online Appendix provides the values of  $\text{mineval}(\Lambda)$  for  $n = \{1, 2, 3\}$  endogenous variables and  $K_2 = n + 1, \dots, 30$  (bias) and  $K_2 = n, \dots, 30$  (size distortion), corresponding to a fine grid of bias and size distortion. Following Stock and Yogo (2005), we calculate the size distortion assuming the Wald test on  $\beta$  has a nominal level of 5%. Using the MATLAB code that we provide, the simulations can be performed at a variety of nominal levels.

Our confidence set provides guidance to researchers on the appropriateness of the instruments they choose for their analysis by constructing a confidence set for the instrument strength, either in terms of bias or size distortion. Our first proposed method, described in Section 3.1.1, applies to the linear IV model with  $n = 1$  endoge-

Figure 2: Bias and size distortion of TSLS estimator as a function of mineval( $\Lambda$ ) ( $n = 1$ )



*Note:* The figures display the bias of the TSLS estimator (left panel) and the size distortion of the corresponding Wald test at the 5% nominal level (right panel) for  $n = 1$  endogenous regressor, and  $K_2$  instruments. The bias values for  $K_2 = 2$  were calculated using the method by Skeels and Windmeijer (2016), while in the remaining cases we followed the simulation approach of Stock and Yogo (2005).

nous regressor. It delivers confidence intervals which are reasonably short and very close to their nominal coverage levels, as we will demonstrate later in the Monte Carlo simulations of [Section 4](#). The second method, described in [Section 3.1.2](#), is generally applicable for any number of endogenous variables ( $n \geq 1$ ), but usually provides more conservative confidence intervals.

### 3.1.1 The Case of One Endogenous Regressor ( $n = 1$ )

The starting point of our proposed confidence interval is the asymptotic distribution of the OLS estimator of  $\Pi$  in [eq. \(14\)](#). Under Assumptions  $L_\Pi$  and  $M$ , the asymptotic distribution of  $\hat{\Pi}_T$  is given by

$$\sqrt{T}\hat{\Pi}_T \xrightarrow{d} \mathcal{N}(C, \sigma_{VV}\Omega^{-1}), \quad (17)$$

which by Slutsky's theorem implies that

$$m_T \equiv \hat{\Omega}^{1/2}\hat{\sigma}_{VV}^{-1/2}\sqrt{T}\hat{\Pi}_T \xrightarrow{d} \mathcal{N}(\Omega^{1/2}C\sigma_{VV}^{-1/2}, I_{K_2}), \quad (18)$$

which in turn leads to

$$f_T \equiv m_T' m_T \xrightarrow{d} \chi_{K_2}^2 \left( K_2 \mu_{K_2}^2 \right), \quad (19)$$

that is  $f_T$  asymptotically follows a non-central chi-squared distribution with  $K_2$  degrees of freedom and non-centrality parameter  $K_2 \mu_{K_2}^2$ . By obtaining a confidence set for  $\mu_{K_2}^2$

and using a projection argument, we can construct an asymptotically valid confidence interval for the bias and the size distortion, as they depend only on  $\mu_{K_2}^2$ , through  $b(\mu_{K_2}^2; n, K_2)$  and  $s(\mu_{K_2}^2; n, K_2)$ , respectively.

Kent and Hainsworth (1995) suggested several confidence intervals for the non-centrality parameter of a chi-squared distribution. Based on their recommendation, we used their proposed ‘‘symmetric range’’ confidence interval. Let  $F_{K_2}(x, K_2\mu_{K_2}^2)$  denote the cumulative distribution function (CDF) of the non-central chi-squared distribution with  $K_2$  degrees of freedom and non-centrality parameter  $K_2\mu_{K_2}^2$  evaluated at  $x$ , and let  $F_{K_2}^{-1}(q, K_2\mu_{K_2}^2)$  denote the corresponding quantile function evaluated at  $q$ . Then the following algorithm leads to  $(1 - \alpha)$  level asymptotic confidence intervals for  $\mu_{K_2}^2$ .

1. Lower bound: If  $\sqrt{f_T} \leq \sqrt{F_{K_2}^{-1}(1 - \alpha, 0)}$ , then set  $l_{1-\alpha}^{\mu_{K_2}^2} = 0$ . Else, solve the equation  $F_{K_2}(f_T, (\sqrt{f_T} - b)^2) - F_{K_2}((\max\{\sqrt{f_T} - 2b, 0\})^2, (\sqrt{f_T} - b)^2) = (1 - \alpha)$  for  $b$ , where  $0 < b < \sqrt{f_T}$ , call the solution  $b^*$ , and set  $l_{1-\alpha}^{\mu_{K_2}^2} = (\sqrt{f_T} - b^*)^2 / K_2$ .
2. Upper bound: Solve the equation  $F_{K_2}((\sqrt{f_T} + 2b)^2, (\sqrt{f_T} + b)^2) - F_{K_2}(f_T, (\sqrt{f_T} + b)^2) = (1 - \alpha)$  for  $b$ , where  $b > 0$ , call the solution  $b^{**}$ . Then set  $u_{1-\alpha}^{\mu_{K_2}^2} = (\sqrt{f_T} + b^{**})^2 / K_2$ .

Then the interval given by  $CI_{1-\alpha}^{\mu_{K_2}^2} \equiv [l_{1-\alpha}^{\mu_{K_2}^2}, u_{1-\alpha}^{\mu_{K_2}^2}]$  is a  $(1 - \alpha)$  level asymptotic confidence interval for  $\mu_{K_2}^2$ . Let us define

$$l_{1-\alpha}^b \equiv b(u_{1-\alpha}^{\mu_{K_2}^2}; n, K_2) \quad u_{1-\alpha}^b \equiv b(l_{1-\alpha}^{\mu_{K_2}^2}; n, K_2), \quad (20)$$

$$l_{1-\alpha}^s \equiv s(u_{1-\alpha}^{\mu_{K_2}^2}; n, K_2) \quad u_{1-\alpha}^s \equiv s(l_{1-\alpha}^{\mu_{K_2}^2}; n, K_2), \quad (21)$$

which constitute the endpoints of the  $(1 - \alpha)$  level asymptotic confidence intervals for bias (eq. (20)) and size distortion (eq. (21)), as summarized in [Proposition 1](#).

**Proposition 1** (Confidence interval validity for  $n = 1$  endogenous regressor): Under Assumptions  $L_\Pi$  and  $M$ ,  $CI_{1-\alpha}^{\mu_{K_2}^2}$  is an asymptotically valid  $(1 - \alpha)$  level confidence interval for  $\mu_{K_2}^2$ , that is,

$$\lim_{T \rightarrow \infty} P \left( \mu_{K_2}^2 \in CI_{1-\alpha}^{\mu_{K_2}^2} \right) = 1 - \alpha. \quad (22)$$

Furthermore,  $[l_{1-\alpha}^b, u_{1-\alpha}^b]$  and  $[l_{1-\alpha}^s, u_{1-\alpha}^s]$  are  $(1 - \alpha)$  level asymptotic confidence

intervals for the bias and size distortion, respectively, formally:

$$\lim_{T \rightarrow \infty} P \left( b \left( \mu_{K_2}^2; n, K_2 \right) \in [l_{1-\alpha}^b, u_{1-\alpha}^b] \right) = 1 - \alpha, \quad (23)$$

$$\lim_{T \rightarrow \infty} P \left( s \left( \mu_{K_2}^2; n, K_2 \right) \in [l_{1-\alpha}^s, u_{1-\alpha}^s] \right) \geq 1 - \alpha. \quad (24)$$

*Proof.* See Section A of the Online Appendix.  $\square$

**Remark 1.** Skeels and Windmeijer (2016) show that, in the case of  $n = 1$  endogenous regressor, the bias  $b \left( \mu_{K_2}^2; n, K_2 \right)$  is a strictly decreasing continuous function of  $\mu_{K_2}^2$  (see their Theorem B.2). If  $s \left( \mu_{K_2}^2; n, K_2 \right)$  is *strictly* decreasing as well (as Stock and Yogo's (2005) simulations strongly suggest), then the corresponding asymptotic confidence interval will not be conservative (the weak inequality in eq. (24) will become an equality).

### 3.1.2 The General Case of Potentially Multiple Endogenous regressors ( $n \geq 1$ )

In the general case of  $n \geq 1$  endogenous regressors, similarly to the previously discussed special case of  $n = 1$ , our proposed confidence interval builds on the asymptotic distribution of the OLS estimator of  $\Pi$  in eq. (14), denoted by  $\hat{\Pi}_T$ :

$$\sqrt{T} \left( \hat{\Pi}_T - \Pi \right) = \left( T^{-1} Z^{\perp \prime} Z^{\perp} \right)^{-1} T^{-1/2} Z^{\perp \prime} V^{\perp}, \quad (25)$$

$$\sqrt{T} \text{vec} \left( \hat{\Pi}_T - \Pi \right) \xrightarrow{d} \mathcal{N} \left( 0, \Sigma_{VV} \otimes \Omega^{-1} \right), \quad (26)$$

$$\text{vec} \left( \hat{C} - C \right) \xrightarrow{d} \mathcal{N} \left( 0, \Sigma_{VV} \otimes \Omega^{-1} \right), \quad (27)$$

where eq. (26) follows directly from eq. (25) and Assumption M, and in eq. (27) we used  $\Pi = \Pi_T = C/\sqrt{T}$  and  $\hat{C} \equiv \hat{\Pi}_T \sqrt{T}$ . While  $\hat{C}$  is an inconsistent estimator of  $C$ , for our purposes the asymptotic normality result of eq. (27) is sufficient. Note that  $\left[ \text{vec} \left( \hat{C} - C \right) \right]' \left[ \Sigma_{VV} \otimes \Omega^{-1} \right]^{-1} \left[ \text{vec} \left( \hat{C} - C \right) \right] \xrightarrow{d} \chi_{nK_2}^2$ . By using  $\hat{\Sigma}_{VV} \xrightarrow{p} \Sigma_{VV}$  and  $\hat{\Omega} \equiv Z^{\perp \prime} Z^{\perp} / T \xrightarrow{p} \Omega$ , we obtain the distribution of the Wald statistic,  $\mathcal{W}(C)$ :

$$\mathcal{W}(C) \equiv \left[ \text{vec} \left( \hat{C} - C \right) \right]' \left[ \hat{\Sigma}_{VV} \otimes \hat{\Omega}^{-1} \right]^{-1} \left[ \text{vec} \left( \hat{C} - C \right) \right] \xrightarrow{d} \chi_{nK_2}^2. \quad (28)$$

By taking the  $(1 - \alpha)$  quantile of the  $\chi_{nK_2}^2$  distribution (denoted by  $\chi_{nK_2, 1-\alpha}^2$ ), the Wald statistic  $\mathcal{W}(C)$  can be inverted to obtain an asymptotically valid  $(1 - \alpha)$  level confidence set for  $C$ , which is formally defined as

$$\text{CI}_{1-\alpha}^C \equiv \left\{ \forall \tilde{C} \in \mathbb{R}^{K_2 \times n} : \mathcal{W} \left( \tilde{C} \right) \leq \chi_{nK_2, 1-\alpha}^2 \right\}. \quad (29)$$

Note that  $\text{CI}_{1-\alpha}^{\text{C}}$  is compact and non-empty by construction. Recall the definition of  $\Lambda$  in eq. (15) and define

$$\tilde{\Lambda}(\tilde{\text{C}}) \equiv \frac{1}{K_2} \hat{\Sigma}_{VV}^{-1/2'} \tilde{\text{C}}' \hat{\Omega} \tilde{\text{C}} \hat{\Sigma}_{VV}^{-1/2},$$

which is a continuous function of  $\tilde{\text{C}}$  and of the consistent estimates of  $\Sigma_{VV}$  and  $\Omega$ . Let us define

$$L_{1-\alpha}^{\Lambda} \equiv \min_{\tilde{\text{C}} \in \text{CI}_{1-\alpha}^{\text{C}}} \text{mineval}(\tilde{\Lambda}(\tilde{\text{C}})) \quad U_{1-\alpha}^{\Lambda} \equiv \max_{\tilde{\text{C}} \in \text{CI}_{1-\alpha}^{\text{C}}} \text{mineval}(\tilde{\Lambda}(\tilde{\text{C}})). \quad (30)$$

Then, following a projection argument (see e.g. Dufour (1997)), a  $(1 - \alpha)$  level asymptotic confidence interval for  $\text{mineval}(\Lambda)$  is given by

$$\text{CI}_{1-\alpha}^{\Lambda} \equiv [L_{1-\alpha}^{\Lambda}, U_{1-\alpha}^{\Lambda}]. \quad (31)$$

Furthermore, let us define

$$L_{1-\alpha}^b \equiv b(U_{1-\alpha}^{\Lambda}; n, K_2) \quad U_{1-\alpha}^b \equiv b(L_{1-\alpha}^{\Lambda}; n, K_2), \quad (32)$$

$$L_{1-\alpha}^s \equiv s(U_{1-\alpha}^{\Lambda}; n, K_2) \quad U_{1-\alpha}^s \equiv s(L_{1-\alpha}^{\Lambda}; n, K_2), \quad (33)$$

which constitute the endpoints of the  $(1 - \alpha)$  level asymptotic confidence intervals for bias (eq. (32)) and size distortion (eq. (33)), as summarized in Proposition 2.

**Proposition 2** (Confidence interval validity for general  $n \geq 1$ ): Under Assumptions  $L_{\Pi}$  and  $M$ ,  $\text{CI}_{1-\alpha}^{\Lambda}$  is an asymptotically valid  $(1 - \alpha)$  level confidence interval for  $\text{mineval}(\Lambda)$ , that is,

$$\lim_{T \rightarrow \infty} P \left( \text{mineval}(\Lambda) \in \text{CI}_{1-\alpha}^{\Lambda} \right) \geq 1 - \alpha. \quad (34)$$

Furthermore,  $[L_{1-\alpha}^b, U_{1-\alpha}^b]$  and  $[L_{1-\alpha}^s, U_{1-\alpha}^s]$  are  $(1 - \alpha)$  level asymptotic confidence intervals for the bias and size distortion, respectively, formally:

$$\lim_{T \rightarrow \infty} P \left( b(\text{mineval}(\Lambda); n, K_2) \in [L_{1-\alpha}^b, U_{1-\alpha}^b] \right) \geq 1 - \alpha, \quad (35)$$

$$\lim_{T \rightarrow \infty} P \left( s(\text{mineval}(\Lambda); n, K_2) \in [L_{1-\alpha}^s, U_{1-\alpha}^s] \right) \geq 1 - \alpha. \quad (36)$$

*Proof.* See Section A of the Online Appendix.  $\square$

**Remark 2.** Note that, as  $\tilde{\Lambda}(\tilde{\text{C}})$  is not a one-to-one function of  $\tilde{\text{C}}$  in general, our proposed confidence interval is conservative.

**Remark 3.** When there is only one endogenous regressor ( $n = 1$ ), then the Karush–Kuhn–Tucker conditions provide an analytical solution to eq. (30), and hence to eqs. (32)

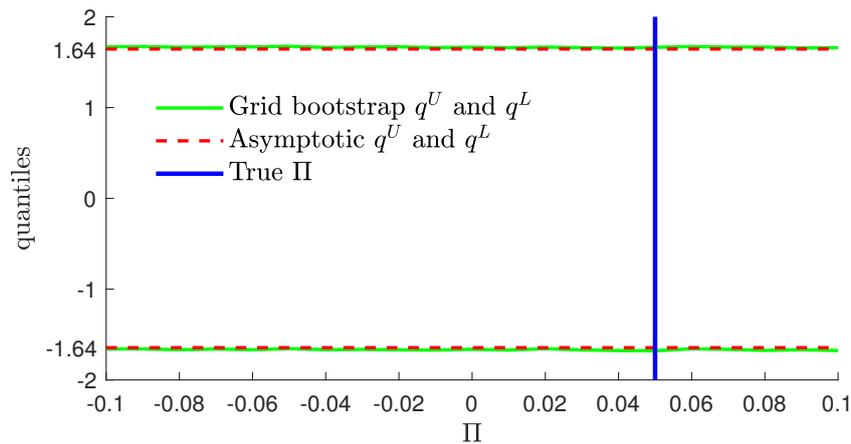
and (33) (we thank an anonymous referee for pointing this out). Define  $d_T \equiv \hat{\sigma}_{VV}^{-1} \hat{C}' \hat{\Omega} \hat{C}$ . The upper bound is given by  $U_{1-\alpha}^\Lambda = K_2^{-1} \left( \sqrt{\chi_{K_2, 1-\alpha}^2} + \sqrt{d_T} \right)^2$ . If  $d_T \geq \chi_{K_2, 1-\alpha}^2$ , then  $L_{1-\alpha}^\Lambda = K_2^{-1} \left( \sqrt{d_T} - \sqrt{\chi_{K_2, 1-\alpha}^2} \right)^2$ , while if  $d_T < \chi_{K_2, 1-\alpha}^2$ , then  $L_{1-\alpha}^\Lambda = 0$  (see Section A of the Online Appendix). However, for a general  $n > 1$ , the lower and upper bounds of the proposed confidence interval must be calculated numerically: we use MATLAB's `fmincon` function to calculate the bounds of the confidence intervals, because the objective function and the constraint are both smooth functions.

Our proposed procedure is an alternative to that of Stock and Yogo (2005). That procedure tests whether the instruments are strong enough either in terms of not leading to an estimator of  $\beta$  more biased than a pre-specified tolerance, or controlling that the Wald test on  $\beta$  does not display higher size distortion than a threshold. Their theory builds on the asymptotic distribution of  $g_{\min}$ . However, their method cannot provide a confidence set for the bias of the TSLS estimator or the size distortion of the corresponding Wald test: that is, researchers do not know *how* weak or strong their instruments are. Our proposed method is specifically designed to provide researchers with such a confidence interval, using a confidence interval of mineval ( $\Lambda$ ), and its relationship with the bias and size distortion of IV estimators.

Alternatively, a uniformly valid confidence interval can be obtained using Hansen's (1999) grid bootstrap. Note that, however, Hansen's (1999) bootstrap is computationally intensive and difficult to implement in multivariate cases. In the following example, we show that, for the case of weak instruments, our procedure (based on asymptotic normality) and Hansen's (1999) deliver the same confidence interval for the strength of identification; our approach, however, is computationally much less intensive.

**Example 1.** *To illustrate the relationship between our asymptotic normal approximation and Hansen's (1999) grid bootstrap, consider a Monte Carlo simulation study. Let us specify  $Y = Z\Pi + X\Phi + V$ , where  $Y, Z, X, V$  are  $(T \times 1)$  vectors such that  $(Z_t, V_t)' \sim \text{iid } \mathcal{N}(0, I_2)$ ,  $\Pi = \Pi_T = C/\sqrt{T}$ ,  $C = 0.5$ ,  $T = 100$ ,  $X_t = 1$  and  $\Phi = 1$ . Hansen's (1999) grid bootstrap is uniformly asymptotically valid in the presence of a weak instrument. If our asymptotic normal approximation is a good approximation of the grid bootstrap quantiles, the 5th and 95th quantiles of the  $t$ -statistic obtained using Hansen's (1999) grid bootstrap are straight lines (i.e. independent of  $\Pi$ ) and equal to  $\pm 1.64$ . Using the grid bootstrap, we can simulate the distribution of the usual  $t$ -statistic testing the null hypothesis of  $\Pi = \Pi_0$  at each point  $\Pi_0$  on a fine grid  $A_G$ , which we specify as ranging from  $-0.1$  to  $0.1$ , with increments of  $0.01$ . At each point on  $A_G$ , we simulate the distribution of the  $t$ -statistic using  $B = 999$  replications and resampling the estimated residuals with replacement; then we estimate the 5th and 95th percentiles ( $q^L$  and  $q^U$ ) of the simulated distribution. Figure 3 depicts the means of  $q^L$  and  $q^U$  at each point on the grid  $A_G$  across 200 replications, confirming that the simulated quantiles of the  $t$ -statistic are virtually indistinguishable from their asymptotic counterparts ( $\pm 1.64$ ).*

Figure 3: The grid bootstrap and asymptotic quantiles of the  $t$ -statistic



Note: The figure reports the 5% and 95% quantiles of the grid bootstrap (solid horizontal lines) as well as the 5% and 95% quantiles of the asymptotic normal approximation on which our projection method relies (dashed horizontal lines), both as a function of  $\Pi$ .

**Example 2.** To illustrate our methodology in an empirical setting, consider Angrist and Krueger's (1991) problem of estimating the returns to education. Angrist and Krueger (1991) estimate the effects of educational attainment on wages, resolving the endogeneity problem using the quarter-of-birth interacted with the year-of-birth as IVs. As Bound et al. (1995) noted, the instruments are only weakly correlated with educational attainment, causing a potential weak instrument problem. Table 1 reports the confidence intervals for bias and size distortions. The first column reports results for the specification in Table V, column 8 in Angrist and Krueger (1991). The TSLS estimate equals 0.060 (with a standard error of 0.029), and the Stock and Yogo (2005)  $F$ -statistic implies that the instruments are weak in terms of bias and size distortion as well. That is, at the 5% significance level one cannot reject the null hypothesis that asymptotically the bias of the TSLS estimator is at most 5% (or even 10%) of the bias of the OLS estimator in the worst case (the worst case corresponds to the biggest relative bias over all possible degrees of simultaneity between the structural and the first-stage errors). Similarly, a researcher cannot reject the null hypothesis that when performing a Wald test on  $\beta$  at the 5% nominal level, asymptotically in the worst case (interpreted as before) he or she would be performing a test which in fact has 5% or 10% larger size than advertised. Our 95% confidence intervals agree with this, no matter whether they are calculated with the projection method or the non-central  $\chi^2$  approximation. The second column reports results for the specification considered by Bound et al. (1995, Table 1, column 2), which includes a smaller number of instruments (only quarter-of-birth). The TSLS estimate is 0.142 (with a standard error of 0.033). The  $F$ -statistic is just below the 5% critical value for bias, and well below the critical value for 5% size distortion. While the Stock and Yogo (2005) test implies weak instruments, a researcher might have ambiguous thoughts about classifying these instruments as weak, as for example the critical value corresponding to 10% bias is 9.08. Indeed, our non-central  $\chi^2$  - based confidence intervals suggest bias between 1.4% and 5.4%, and size distortion between 3.0% and 9.7%, which an applied researcher might be comfortable with.

Table 1: Estimating the returns to education: confidence intervals

	<i>Angrist and Krueger (1991)</i> ( $K_2 = 28$ )	<i>Bound et al. (1995)</i> ( $K_2 = 3$ )
TOLS estimate (standard error)	0.060 (0.029)	0.142 (0.033)
95% Confidence intervals for bias		
Projection method	[0.132; 0.997]	[0.012; 0.087]
Non-central $\chi^2$	[0.223; 0.914]	[0.014; 0.054]
95% Confidence intervals for size distortion		
Projection method	[0.532; 0.950]	[0.024; 0.140]
Non-central $\chi^2$	[0.777; 0.950]	[0.030; 0.097]
<i>F</i> -statistic	1.61	13.49
Critical value (5% bias)	21.42	13.91
Critical value (10% bias)	11.34	<b>9.08</b>
Critical value (5% size distortion)	81.40	22.30
Critical value (10% size distortion)	42.37	<b>12.83</b>

*Note:* The upper panel reports confidence intervals for bias and size distortion in the Angrist and Krueger (1991) and the Bound et al. (1995) returns to education regressions. The lower panel shows the *F*-statistics and the corresponding critical values (at the 5% significance level) for bias and size distortion (nominal level of Wald test is 5%) following Stock and Yogo (2005). Critical values in bold correspond to strong instruments according to the specific threshold.

### 3.2 The Heteroskedastic/Autocorrelated Linear IV Model

The assumption of homoskedastic errors used in the previous section may be restrictive in a number of applications. In those cases, applying either the Stock and Yogo (2005) test or our proposed confidence interval could lead to incorrect inference on the instrument strength. As a solution to this problem, Montiel Olea and Pflueger (2013) propose a measure of the strength of instruments which applies to general (heteroskedastic, autocorrelated or clustered) errors, albeit the theory has been developed for the case of  $n = 1$  endogenous regressor. They consider the TOLS and the limited information maximum likelihood (LIML) estimators. For simplicity and due to its popularity, in our paper we focus on the TOLS estimator.

Following Montiel Olea and Pflueger (2013), consider the linear IV model in its reduced form:

$$y^\perp = Z^\perp \Pi \beta + v_1, \quad (37)$$

$$Y^\perp = Z^\perp \Pi + v_2, \quad (38)$$

where eq. (37) is the structural equation of interest in reduced form, while eq. (38) is the first stage equation linking the endogenous regressor  $Y^\perp$  with the instruments  $Z^\perp$  (both projected on the exogenous variables). Both  $y^\perp$  and  $Y^\perp$  are  $(T \times 1)$  vectors,  $Z^\perp$  is a  $(T \times K_2)$  matrix of instruments,  $\beta$  is a scalar coefficient,  $\Pi$  is a  $(K_2 \times 1)$  vector

of coefficients, while  $v_1 \equiv V^\perp \beta + u^\perp$  and  $v_2 \equiv V^\perp$  are  $(T \times 1)$  vectors of errors. Furthermore, in this section,  $Z^\perp$  is orthogonalized such that  $Z^{\perp'} Z^\perp / T = I_{K_2}$ .

Montiel Olea and Pflueger (2013) adopt Assumption  $L_\Pi$  of SSY to model weak instruments, but considerably weaken their moment assumptions as follows:

**Assumption HL** *The following limits hold as  $T \rightarrow \infty$ :*

$$(a) \begin{pmatrix} T^{-1/2} Z^{\perp'} v_1 \\ T^{-1/2} Z^{\perp'} v_2 \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, W) \text{ for some positive definite } W = \begin{pmatrix} W_1 & W_{12} \\ W'_{12} & W_2 \end{pmatrix}, \text{ where} \\ \text{the sub-matrices of } W \text{ are all } (K_2 \times K_2) \text{ square matrices;}$$

$$(b) [v_1 \ v_2]' [v_1 \ v_2] / T \xrightarrow{p} \kappa \text{ for some positive definite } \kappa;$$

$$(c) \text{ There exists a sequence of positive definite estimates } \widehat{W}, \text{ measurable with respect to} \\ \{y_t^\perp, Y_t^\perp, Z_t^\perp\}_{t=1}^T, \text{ such that } \widehat{W} \xrightarrow{p} W.$$

Unlike Assumption M of SSY, these high level assumptions do not restrict  $W$  to take the form of  $\kappa \otimes I_{K_2}$ , and therefore they can encompass a wide range of error structures, including heteroskedastic, autocorrelated or clustered (in panel data) error terms.

Montiel Olea and Pflueger (2013) formulate their notion of weak instruments in terms of the Nagar (1959) bias, which is defined as

$$N_{\text{TSLs}}(\beta, C, W) \equiv \mu^{-2} \frac{\text{tr}(S_{12})}{\text{tr}(S_2)} \left[ 1 - 2 \frac{C'_0 S_{12} C_0}{\text{tr}(S_{12})} \right], \quad (39)$$

where  $C = \|C\|C_0$ ,  $\mu^2 \equiv \|C\|^2 / \text{tr}(W_2)$ ,  $S_1 \equiv W_1 - 2\beta W_{12} + \beta^2 W_2$ ,  $S_{12} \equiv W_{12} - \beta W_2$ , and  $S_2 \equiv W_2$ . Note that  $\mu^2$  can be thought of as the analog of the concentration parameter  $\mu_{K_2}^2$  defined in Section 3.1. The Nagar bias is the expected value of the first three terms in the Taylor expansion of the asymptotic distribution of the TSLs estimator under weak instrument asymptotics (in the case of irrelevant instruments, corresponding to  $C = 0$ , we define the Nagar bias as either  $+\infty$  or  $-\infty$ ). Furthermore, they define the benchmark ‘‘worst-case’’ bias as  $\text{BM}(\beta, W) \equiv \sqrt{\text{tr}(S_1) / \text{tr}(S_2)}$ , which is intuitively related to the approximate bias of the TSLs estimator when the instruments are uninformative and the first-stage and second-stage errors are perfectly correlated (see Remark 4 on p. 362 in Montiel Olea and Pflueger, 2013). Then, for a given threshold  $\tau \in [0, 1]$  (specified by the researcher) they define the weak instrument set as

$$\mu^2 \in \mathbb{R}_+ : \sup_{\beta \in \mathbb{R}, C_0 \in \mathcal{S}^{K_2-1}} \frac{|N_{\text{TSLs}}(\beta, \mu \sqrt{\text{tr}(W_2)} C_0, W)|}{\text{BM}(\beta, W)} > \tau, \quad (40)$$

where  $\mathcal{S}^{K_2-1}$  is the  $K_2 - 1$  dimensional unit sphere. That is, the instruments are weak if the Nagar bias exceeds a fraction  $\tau$  of the benchmark bias  $\text{BM}(\beta, W)$  for at least some value of the structural parameter  $\beta$  and some direction of the first-stage coefficients

C<sub>0</sub>. Montiel Olea and Pflueger (2013) propose the so-called *effective first-stage F-statistic* to test the null hypothesis of weak instruments:

$$\widehat{F}_{\text{eff}} \equiv \frac{Y^{\perp'}(Z^{\perp}Z^{\perp'}/T)Y^{\perp}}{\text{tr}(\widehat{W}_2)}. \quad (41)$$

However, their procedure cannot guide researchers on *how* weak or strong their instruments are. On the other hand, our proposed methodology allows researchers to go beyond hypothesis testing by providing an asymptotic confidence interval for the Nagar bias defined in eq. (39). Therefore our procedure has the additional advantage of providing information on the bias of the TSLS estimator directly, without the need to relate it to the worst-case benchmark bias  $\text{BM}(\beta, W)$ . However, an additional complication arises due to the fact that the Nagar bias depends on the structural parameter  $\beta$ , which is not consistently estimable under weak instrument asymptotics.

In what follows, we explain how we can still provide a confidence interval for the Nagar bias. Our proposed confidence interval for the Nagar bias is constructed by combining two ideas: the method of obtaining a confidence set for  $C$  in Section 3.1 and the method of obtaining a confidence set for  $\beta$ . Consider the following compact expression of eqs. (37) and (38) (used by e.g. Andrews, Moreira and Stock (2006)):

$$\widetilde{Y} = Z^{\perp}\Pi a' + \widetilde{v}, \quad (42)$$

where  $\widetilde{Y} \equiv [y^{\perp}, Y^{\perp}]$ ,  $a \equiv (\beta, 1)'$  and  $\widetilde{v} \equiv [v_1, v_2]$ . Note that the coefficient matrix  $\Pi a'$  has an interesting structure: its first column is  $\Pi\beta$ , while its second column is  $\Pi$ . Let us define its vectorized version as  $\Gamma \equiv [\Pi'\beta, \Pi']'$ , then vectorize eq. (42):

$$\text{vec}(\widetilde{Y}) = (I_2 \otimes Z^{\perp})\Gamma + \text{vec}(\widetilde{v}). \quad (43)$$

Consider the asymptotic distribution of the OLS estimator of  $\Gamma$  in eq. (43):

$$\sqrt{T}(\widehat{\Gamma} - \Gamma) = \begin{bmatrix} T^{-1/2}Z^{\perp'v_1} \\ T^{-1/2}Z^{\perp'v_2} \end{bmatrix} = \begin{bmatrix} \widehat{\psi} - C\beta \\ \widehat{C} - C \end{bmatrix} \xrightarrow{d} \mathcal{N}(0, W), \quad (44)$$

where  $\widehat{\psi}$  is  $\sqrt{T}$  times the OLS estimator of  $\Pi_T\beta$  in the structural equation,  $\widehat{C} = \widehat{\Pi}_T\sqrt{T}$  as in Section 3.1.2, and we used the normalization  $Z^{\perp'Z^{\perp}}/T = I_{K_2}$  and Assumptions HL and L<sub>II</sub>. Furthermore, by Slutsky's theorem and part (c) of Assumption HL, the Wald statistic asymptotically follows a chi-squared distribution with  $2K_2$  degrees of freedom, formally

$$\mathcal{W}(C, \beta) \equiv \begin{bmatrix} \widehat{\psi} - C\beta \\ \widehat{C} - C \end{bmatrix}' \widehat{W}^{-1} \begin{bmatrix} \widehat{\psi} - C\beta \\ \widehat{C} - C \end{bmatrix} \xrightarrow{d} \chi_{2K_2}^2. \quad (45)$$

Analogously to the procedure in [Section 3.1](#), by taking the  $(1 - \alpha)$  quantile of the  $\chi_{2K_2}^2$  distribution (denoted by  $\chi_{2K_2, 1-\alpha}^2$ ), the Wald statistic  $\mathcal{W}(C, \beta)$  can be inverted to obtain an asymptotically valid  $(1 - \alpha)$  level *joint* confidence set for  $C$  and  $\beta$ , formally:

$$\text{CI}_{1-\alpha}^{C, \beta} \equiv \left\{ \forall (\tilde{C}, \tilde{\beta}) \in \mathbb{R}^{K_2+1} : \mathcal{W}(\tilde{C}, \tilde{\beta}) \leq \chi_{2K_2, 1-\alpha}^2 \right\}. \quad (46)$$

Note that if  $\tilde{C} = 0$  is in the confidence set  $\text{CI}_{1-\alpha}^{C, \beta}$ , then the confidence set for  $\beta$  is unbounded, which is in line with the findings of Dufour (1997). If this is the case, then it suggests that the instruments are very weak indeed, and  $\beta$  might not be identified. Therefore when  $\tilde{C} = 0 \in \text{CI}_{1-\alpha}^{C, \beta}$ , then we take  $[-\infty, +\infty]$  as our confidence set for the Nagar bias. Another peculiar case is when the confidence set  $\text{CI}_{1-\alpha}^{C, \beta}$  is empty, which can happen when a confidence set is based on the inversion principle. This situation indicates that the data rejects the model, pointing to the violation of the exclusion restriction. In this case, we take the empty set (denoted by  $\emptyset$ ) as the confidence set for the Nagar bias. We proceed to describe our proposed confidence intervals, keeping in mind these special cases.

To construct a confidence interval for the Nagar bias, let us define the Nagar bias as a function of the parameters  $(\tilde{C}, \tilde{\beta})$  and the consistent estimate  $\hat{W}$ :

$$\tilde{N}_{\text{TSLs}}(\tilde{\beta}, \tilde{C}, \hat{W}) \equiv \tilde{\mu}^{-2} \frac{\text{tr}(\tilde{S}_{12})}{\text{tr}(\tilde{S}_2)} \left[ 1 - 2 \frac{\tilde{C}'_0 \tilde{S}_{12} \tilde{C}_0}{\text{tr}(\tilde{S}_{12})} \right], \quad (47)$$

where  $\tilde{C} = \|\tilde{C}\| \tilde{C}_0$ ,  $\tilde{\mu}^2 \equiv \|\tilde{C}\|^2 / \text{tr}(\hat{W}_2)$ ,  $\tilde{S}_1 \equiv \hat{W}_1 - 2\tilde{\beta} \hat{W}_{12} + \tilde{\beta}^2 \hat{W}_2$ ,  $\tilde{S}_{12} \equiv \hat{W}_{12} - \tilde{\beta} \hat{W}_2$ , and  $\tilde{S}_2 \equiv \hat{W}_2$ .

Let us define  $L_{1-\alpha}^N \equiv \min_{(\tilde{C}, \tilde{\beta}) \in \text{CI}_{1-\alpha}^{C, \beta}} \tilde{N}_{\text{TSLs}}(\tilde{\beta}, \tilde{C}, \hat{W})$  and  $U_{1-\alpha}^N \equiv \max_{(\tilde{C}, \tilde{\beta}) \in \text{CI}_{1-\alpha}^{C, \beta}} \tilde{N}_{\text{TSLs}}(\tilde{\beta}, \tilde{C}, \hat{W})$ .

Our proposed  $(1 - \alpha)$  level asymptotic confidence interval for  $N_{\text{TSLs}}(\beta, C, W)$  is

$$\text{CI}_{1-\alpha}^{N_{\text{TSLs}}} = \begin{cases} [L_{1-\alpha}^N, U_{1-\alpha}^N] & \text{if } \text{CI}_{1-\alpha}^{C, \beta} \neq \emptyset \text{ and } \tilde{C} = 0 \notin \text{CI}_{1-\alpha}^{C, \beta}, \\ [-\infty, +\infty] & \text{if } \tilde{C} = 0 \in \text{CI}_{1-\alpha}^{C, \beta}, \\ \emptyset & \text{if } \text{CI}_{1-\alpha}^{C, \beta} = \emptyset. \end{cases} \quad (48)$$

We summarize our results in the following proposition.

**Proposition 3** (Confidence interval validity under general assumptions): Under Assumptions  $L_{\Pi}$  and HL,  $\text{CI}_{1-\alpha}^{N_{\text{TSLs}}}$  in [eq. \(48\)](#) is an asymptotically valid confidence interval for the Nagar bias  $N_{\text{TSLs}}(\beta, C, W)$ , that is

$$\lim_{T \rightarrow \infty} \text{P} \left( N_{\text{TSLs}}(\beta, C, W) \in \text{CI}_{1-\alpha}^{N_{\text{TSLs}}} \right) \geq 1 - \alpha. \quad (49)$$

*Proof.* See Section A of the Online Appendix. □

Note that, as  $\tilde{N}_{\text{TSLs}}(\tilde{\beta}, \tilde{C}, \tilde{W})$  is not a one-to-one function of  $(\tilde{C}, \tilde{\beta})$  in general, our proposed confidence interval may be conservative.

### 3.3 Local Projections–IV method

Since the original paper by Jordà (2005), the local projections method has become popular in the macroeconomics literature to estimate impulse response functions, due to its simplicity (both in terms of estimation and inference) and robustness to model misspecification. Its IV variant called local projections–IV (LP–IV) has been used in several recent studies, see for example Jordà et al. (2015) examining the link between financial conditions, mortgage credit and house prices, or Ramey and Zubairy (2018) investigating state-dependent US government spending multipliers. Stock and Watson (2018) provides an overview of the LP–IV econometric framework, which we adopt here.

Consider the  $(n \times 1)$  vector of covariance stationary macroeconomic variables  $Y_t$ , and its structural vector moving average representation  $Y_t = \Theta(L)\varepsilon_t$ , where  $L$  is the lag operator,  $\Theta(L) = \Theta_0 + \Theta_1 L + \Theta_2 L^2 + \dots$ , and  $\Theta_h$  is an  $(n \times m)$  matrix of coefficients. Furthermore,  $\varepsilon_t$  is an  $(m \times 1)$  vector of mutually uncorrelated structural shocks and measurement errors with a positive definite covariance matrix. The coefficients of  $\Theta(L)$  are the structural impulse response functions. Suppose that the researcher is interested in the response of the  $i$ -th endogenous variable at horizon  $h$ ,  $Y_{i,t+h}$ , to a unitary increase in  $\varepsilon_{1,t}$ , and let the  $(i, 1)$  element of  $\Theta_h$  be denoted by  $\Theta_{h,i1}$ . A convenient normalization is  $\Theta_{0,11} = 1$ , that is a unit increase in  $\varepsilon_{1,t}$  leads to a unit increase in  $Y_{1,t}$ . It follows that we can write  $Y_{1,t} = \varepsilon_{1,t} + \{\varepsilon_{2:n,t}, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$ , where  $\varepsilon_{2:n,t} \equiv (\varepsilon_{2,t}, \varepsilon_{3,t}, \dots, \varepsilon_{n,t})'$ , and the shorthand  $\{\cdot\}$  denotes the linear combination of the variables inside the braces. Then the  $h$ -period-ahead impulse response of the  $i$ -th variable  $Y_{i,t+h}$  to a structural shock  $\varepsilon_{1,t}$  is given by the population coefficient in the linear regression

$$Y_{i,t+h} = \Theta_{h,i1} Y_{1,t} + u_{i,t+h}^h, \quad (50)$$

where  $u_{i,t+h}^h = \{\varepsilon_{t+h}, \dots, \varepsilon_{t+1}, \varepsilon_{2:n,t}, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$ . Given the endogeneity of  $Y_{1,t}$ , OLS is inconsistent, but TSLs is consistent if an appropriate set of instrumental variables  $Z_t$  is available. Note that in general,  $u_{i,t+h}^h$  is serially correlated for  $h > 1$  by construction.

As common in the empirical literature, a vector of control variables  $X_t$  can be added to eq. (50), resulting in  $Y_{i,t+h} = \Theta_{h,i1} Y_{1,t} + \gamma_h' X_t + u_{i,t+h}^h$ . After projecting on the control variables, the regression of interest becomes

$$Y_{i,t+h}^\perp = \Theta_{h,i1} Y_{1,t}^\perp + u_{i,t+h}^{h\perp}. \quad (51)$$

As Stock and Watson (2018) note, the control variables can serve two purposes: first, the exogeneity conditions  $E(\varepsilon_{2:n,t}Z_t') = 0$ , and  $E(\varepsilon_{t+j}Z_t') = 0$  for all  $j \neq 0$  might only be satisfied after controlling for  $X_t$ . Second, they can reduce the variance of the IV estimator through reducing the variance of the error term. The exogeneity conditions in the presence of control variables are  $E(\varepsilon_{2:n,t}Z_t^{\perp'}) = 0$ , and  $E(\varepsilon_{t+j}^{\perp}Z_t^{\perp'}) = 0$  for all  $j \neq 0$ . Instrument relevance is given by  $E(\varepsilon_{1,t}^{\perp}Z_t^{\perp}) = \Pi$ . Note that under instrument exogeneity, the instrument relevance condition is equivalent to  $E(Y_{1,t}^{\perp}Z_t^{\perp}) = \Pi$ , which suggests the familiar first stage equation (using the same normalization  $Z_t^{\perp'}Z_t^{\perp}/T = I_{K_2}$  as before, and  $Y_{1,t}^{\perp}$  acting as the endogenous regressor  $Y_t^{\perp}$ ):

$$Y_{1,t}^{\perp} = Z_t^{\perp'}\Pi + v_{2,t}, \quad (52)$$

where  $E(v_{2,t}Z_t^{\perp}) = 0$ . From an IV perspective, the structural equation in its reduced form is given by

$$Y_{i,t+h}^{\perp} = Z_t^{\perp'}\Pi\Theta_{h,i1} + \Theta_{h,i1}v_{2,t} + u_{i,t+h}^{h\perp} = Z_t^{\perp'}\Pi\Theta_{h,i1} + v_{1,t}, \quad (53)$$

where  $Y_{i,t+h}^{\perp}$  corresponds to  $y_t$ ,  $\Theta_{h,i1}$  plays the role of  $\beta$ , and  $\Theta_{h,i1}v_{2,t} + u_{i,t+h}^{h\perp}$  is equivalent to  $v_{1,t}$  in the heteroskedastic/autocorrelated IV model. As Stock and Watson (2018) note, apart from special cases, by construction  $Z_t^{\perp'}v_{1,t}$  and  $Z_t^{\perp'}v_{2,t}$  feature conditional heteroskedasticity and autocorrelation, hence our confidence interval in the previous subsection applies directly to the LP-IV framework under Assumptions L $\Pi$ , HL, instrument exogeneity, and the validity of the structural vector moving average representation described at the beginning of this subsection.

## 4 MONTE CARLO ANALYSIS

In this section, we investigate the performance of the confidence intervals that we proposed in both the homoskedastic, and the heteroskedastic and serially correlated IV model. Throughout, we focus on the empirical coverage rates of our proposed confidence intervals; in the homoskedastic IV model with  $n = 1$  we provide median lengths as well, to compare the projection method to the non-central chi-squared approach. The Online Appendix provides further results, including the median lengths of the confidence intervals. Without loss of generality, in this section we do not include exogenous regressors (thus,  $Y = Y^{\perp}$ ,  $Z = Z^{\perp}$  and  $V = V^{\perp}$ ). The number of Monte Carlo replications is 2000 and the nominal level of the confidence intervals' coverage is  $(1 - \alpha) = 0.90$  in all designs.

## 4.1 The Homoskedastic IV Model

Let the first stage equation be:

$$Y = Z\Pi + V, \quad (54)$$

where  $Y$  is the  $T \times n$  matrix of endogenous variables,  $Z$  is the  $T \times K_2$  matrix of instruments, and  $V$  is the  $T \times n$  matrix of errors. We specify  $V_t \sim iid \mathcal{N}(0, I_n)$  and  $Z_t \sim iid \mathcal{N}(0, I_{K_2})$ , and consider  $n = \{1, 2\}$ , with  $K_2 = \{n + 1, \dots, n + 4\}$  when focusing on bias, and  $K_2 = \{n, \dots, n + 3\}$  when analyzing size distortion. For each pair  $(n, K_2)$ , we consider three values of bias and size distortion: 5%, 10% and 30% (Section C of the Online Appendix contains the values of  $C$  used in the simulations). We consider sample sizes of  $T = \{100, 250, 500, 1000\}$ . When there is only one endogenous regressor ( $n = 1$ ) we constructed the confidence intervals based on both the non-central chi-squared approach of [Proposition 1](#) and the projection method of [Proposition 2](#). Doing so allows us to evaluate the conservativeness of the projection method. Recall that in the homoskedastic model, bias and size distortion do not depend on structural equation parameters, only on the smallest eigenvalue of the concentration matrix,  $\text{mineval}(\Lambda)$ ,  $K_2$ , and  $n$ .

As Panels A of [Tables 2](#) and [3](#) show, the confidence intervals based on the non-central chi-squared approximation display coverage rates very close to the nominal 90% level for a variety of sample sizes and bias/size distortion values. The coverage rates of the confidence intervals based on the projection method are shown in Panels B of [Tables 2](#) and [3](#), calculated using exactly the same simulated data. As we can see, these confidence intervals exhibit over-coverage (as anticipated), which increases in the number of instruments  $K_2$ , and for a given  $K_2$  it is smaller for smaller values of bias/size distortion (*modulo* Monte Carlo error). The intuition behind the former is that the larger the dimension of the vector  $C$ , the “less” one-to-one  $\tilde{\Lambda}(\tilde{C})$  becomes. The latter effect is due to the fact that smaller values of bias/size distortion correspond to larger values of  $C$ , which are further away from the origin, thereby further away from a part of the parameter space where  $\tilde{\Lambda}(\tilde{C})$  is particularly non-invertible.

Panels C and D of [Tables 2](#) and [3](#) illustrate that the median lengths of confidence intervals are slightly larger with the projection method than with the non-central  $\chi^2$  approximation.

[Table 4](#) shows that for  $n = 2$  endogenous variables, our projection method-based confidence intervals are conservative in general.

Overall, our methods perform well across different specifications, even for relatively small samples.

Table 2: Homoskedastic IV model,  $n = 1$  endogenous variable, confidence intervals for TSLS bias  $b$

<b>Panel A. Coverage rates (non-central <math>\chi^2</math>)</b>												
$T \setminus b$	$K_2 = 2$			$K_2 = 3$			$K_2 = 4$			$K_2 = 5$		
	0.3	0.1	0.05	0.3	0.1	0.05	0.3	0.1	0.05	0.3	0.1	0.05
100	0.89	0.88	0.89	0.90	0.87	0.86	0.89	0.86	0.83	0.87	0.83	0.79
250	0.90	0.90	0.90	0.88	0.88	0.88	0.90	0.89	0.87	0.89	0.87	0.85
500	0.90	0.90	0.89	0.91	0.91	0.90	0.90	0.89	0.89	0.91	0.90	0.90
1000	0.90	0.89	0.89	0.90	0.90	0.89	0.90	0.90	0.90	0.91	0.90	0.89

<b>Panel B. Coverage rates (projection method)</b>												
$T \setminus b$	$K_2 = 2$			$K_2 = 3$			$K_2 = 4$			$K_2 = 5$		
	0.3	0.1	0.05	0.3	0.1	0.05	0.3	0.1	0.05	0.3	0.1	0.05
100	0.96	0.96	0.96	0.97	0.97	0.97	0.99	0.99	0.98	0.99	0.99	0.98
250	0.97	0.97	0.97	0.98	0.98	0.98	0.99	0.99	0.99	0.99	0.99	0.99
500	0.97	0.97	0.97	0.98	0.99	0.99	0.99	0.99	0.99	0.99	1.00	1.00
1000	0.98	0.98	0.98	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99

<b>Panel C. Median lengths of confidence intervals (non-central <math>\chi^2</math>)</b>												
$T \setminus b$	$K_2 = 2$			$K_2 = 3$			$K_2 = 4$			$K_2 = 5$		
	0.3	0.1	0.05	0.3	0.1	0.05	0.3	0.1	0.05	0.3	0.1	0.05
100	0.49	0.49	0.47	0.84	0.28	0.09	0.73	0.16	0.05	0.57	0.12	0.04
250	0.49	0.49	0.48	0.83	0.27	0.09	0.72	0.16	0.05	0.56	0.11	0.04
500	0.49	0.49	0.47	0.84	0.29	0.09	0.72	0.16	0.05	0.59	0.12	0.04
1000	0.49	0.49	0.48	0.83	0.28	0.09	0.73	0.16	0.05	0.60	0.12	0.04

<b>Panel D. Median lengths of confidence intervals (projection method)</b>												
$T \setminus b$	$K_2 = 2$			$K_2 = 3$			$K_2 = 4$			$K_2 = 5$		
	0.3	0.1	0.05	0.3	0.1	0.05	0.3	0.1	0.05	0.3	0.1	0.05
100	0.50	0.50	0.50	0.93	0.63	0.22	0.90	0.40	0.12	0.87	0.30	0.09
250	0.50	0.50	0.50	0.93	0.60	0.21	0.90	0.40	0.12	0.87	0.28	0.09
500	0.50	0.50	0.50	0.93	0.64	0.23	0.90	0.39	0.12	0.88	0.30	0.09
1000	0.50	0.50	0.50	0.93	0.62	0.22	0.90	0.40	0.12	0.87	0.30	0.09

Note: Panel A shows the empirical coverage rates of the proposed confidence interval for the TSLS bias  $b$  based on the non-central  $\chi^2$  approximation for different sample sizes  $T$ , values of  $b$ , and number of instruments  $K_2$  for the homoskedastic DGP. Panel B displays analogous results, based on the projection method. Panels C and D report median lengths of the confidence intervals. The number of Monte Carlo simulations is 2000. The nominal coverage level is  $(1 - \alpha) = 0.90$ .

## 4.2 The Heteroskedastic/Autocorrelated IV Model

We consider two DGPs (Data Generating Processes), labeled as DGP 1 and DGP 2, and construct confidence intervals for the Nagar bias defined in eq. (39). DGP 1 is inspired by Montiel Olea and Pflueger (2013, p. 361), and features conditional heteroskedasticity but no autocorrelation, while DGP 2 has both. First we describe DGP 2, and then discuss the restriction under which we obtain DGP 1.

Let  $\tilde{Z}_t = (\tilde{Z}_{1,t}, \dots, \tilde{Z}_{K_2,t})'$ ,  $\epsilon_t \sim iid\mathcal{N}(0, (1 - \rho^2)I_{K_2})$  and  $\tilde{Z}_t = \rho\tilde{Z}_{t-1} + \epsilon_t$ , where

Table 3: Homoskedastic IV model,  $n = 1$  endogenous variable, confidence intervals for size distortion  $s$

<b>Panel A. Coverage rates (non-central <math>\chi^2</math>)</b>												
$T \setminus s$	$K_2 = 1$			$K_2 = 2$			$K_2 = 3$			$K_2 = 4$		
	0.3	0.1	0.05	0.3	0.1	0.05	0.3	0.1	0.05	0.3	0.1	0.05
100	0.91	0.90	0.89	0.89	0.88	0.86	0.89	0.86	0.82	0.89	0.85	0.80
250	0.89	0.89	0.89	0.90	0.89	0.88	0.89	0.88	0.87	0.90	0.88	0.86
500	0.91	0.91	0.90	0.90	0.89	0.89	0.91	0.90	0.89	0.89	0.89	0.88
1000	0.90	0.89	0.90	0.90	0.89	0.89	0.90	0.90	0.89	0.90	0.90	0.89

<b>Panel B. Coverage rates (projection method)</b>												
$T \setminus s$	$K_2 = 1$			$K_2 = 2$			$K_2 = 3$			$K_2 = 4$		
	0.3	0.1	0.05	0.3	0.1	0.05	0.3	0.1	0.05	0.3	0.1	0.05
100	0.95	0.96	0.89	0.96	0.95	0.94	0.97	0.97	0.96	0.99	0.98	0.96
250	0.94	0.94	0.89	0.97	0.96	0.96	0.98	0.98	0.97	0.99	0.99	0.98
500	0.94	0.95	0.90	0.97	0.97	0.96	0.98	0.99	0.98	0.99	0.99	0.99
1000	0.94	0.94	0.90	0.98	0.97	0.97	0.98	0.99	0.99	0.99	0.99	0.99

<b>Panel C. Median lengths of confidence intervals (non-central <math>\chi^2</math>)</b>												
$T \setminus b$	$K_2 = 1$			$K_2 = 2$			$K_2 = 3$			$K_2 = 4$		
	0.3	0.1	0.05	0.3	0.1	0.05	0.3	0.1	0.05	0.3	0.1	0.05
100	0.89	0.89	0.09	0.81	0.21	0.07	0.68	0.14	0.05	0.56	0.11	0.04
250	0.89	0.88	0.09	0.81	0.22	0.07	0.66	0.14	0.05	0.56	0.11	0.04
500	0.89	0.88	0.09	0.81	0.20	0.07	0.70	0.15	0.05	0.56	0.11	0.04
1000	0.89	0.88	0.09	0.82	0.22	0.07	0.68	0.14	0.05	0.57	0.11	0.04

<b>Panel D. Median lengths of confidence intervals (projection method)</b>												
$T \setminus b$	$K_2 = 1$			$K_2 = 2$			$K_2 = 3$			$K_2 = 4$		
	0.3	0.1	0.05	0.3	0.1	0.05	0.3	0.1	0.05	0.3	0.1	0.05
100	0.89	0.89	0.14	0.86	0.38	0.11	0.83	0.30	0.09	0.81	0.25	0.08
250	0.89	0.89	0.15	0.86	0.41	0.11	0.83	0.29	0.09	0.80	0.25	0.08
500	0.89	0.89	0.15	0.86	0.38	0.11	0.83	0.30	0.09	0.81	0.24	0.08
1000	0.89	0.89	0.14	0.86	0.41	0.11	0.83	0.30	0.09	0.81	0.25	0.08

Note: Panel A shows the empirical coverage rates of the proposed confidence interval for size distortion  $s$  (nominal level of Wald test is 5%) based on the non-central  $\chi^2$  approximation for different sample sizes  $T$ , values of  $s$ , and number of instruments  $K_2$  for the homoskedastic DGP. Panel B displays analogous results, based on the projection method. Panels C and D report median lengths of the confidence intervals. The number of Monte Carlo simulations is 2000. The nominal coverage level is  $(1 - \alpha) = 0.90$ .

$\rho$  controls the persistence of the independent autoregressive processes, and we set  $\rho = 0.7$ . The  $(T \times K_2)$  matrix  $\tilde{Z}$  collects the vectors  $\tilde{Z}_t$ . Let  $Z_t$  be the standardized  $\tilde{Z}_t$  in-sample, such that  $Z'Z/T = I_{K_2}$ . That is, let  $Z^{\text{std}}$  be the (column-by-column) demeaned  $\tilde{Z}$  divided by its standard deviation (column-by-column), and  $Q_Z^{\text{std}} = (Z^{\text{std}'Z^{\text{std}}/T)^{-1/2}$ . Then  $Z = Z^{\text{std}}Q_Z^{\text{std}}$ . We specify a moving average process  $u_{2t} = q_t + \theta q_{t-1}$ , where  $q_t \sim iid\mathcal{N}(0, 1)$ , and  $\theta = 0.4$ , and it is independent of  $\tilde{Z}_t$  (and  $Z_t$ ) both contemporaneously and at all leads and lags. Furthermore, let  $b_t \sim iid\mathcal{N}(0, 1)$

Table 4: Homoskedastic IV model,  $n = 2$  endogenous variables, coverage rates for TOLS bias  $b$  and size distortion  $s$

Panel A. Confidence intervals for TOLS bias $b$												
$T \setminus b$	$K_2 = 3$			$K_2 = 4$			$K_2 = 5$			$K_2 = 6$		
	0.3	0.1	0.05	0.3	0.1	0.05	0.3	0.1	0.05	0.3	0.1	0.05
100	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	0.95	1.00	1.00	0.90
250	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	0.96	1.00	1.00	0.90
500	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.97	1.00	1.00	0.92
1000	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	0.96	1.00	1.00	0.92

Panel B. Confidence intervals for size distortion $s$												
$T \setminus s$	$K_2 = 2$			$K_2 = 3$			$K_2 = 4$			$K_2 = 5$		
	0.3	0.1	0.05	0.3	0.1	0.05	0.3	0.1	0.05	0.3	0.1	0.05
100	1.00	1.00	1.00	1.00	1.00	0.95	1.00	1.00	0.87	1.00	0.99	0.74
250	1.00	1.00	1.00	1.00	1.00	0.97	1.00	1.00	0.88	1.00	1.00	0.73
500	1.00	1.00	1.00	1.00	1.00	0.96	1.00	1.00	0.90	1.00	1.00	0.74
1000	1.00	1.00	1.00	1.00	1.00	0.96	1.00	1.00	0.88	1.00	1.00	0.76

Note: Panel A shows the empirical coverage rates of the proposed confidence interval for the TOLS bias  $b$  based on the projection method for different sample sizes  $T$ , values of  $b$ , and number of instruments  $K_2$  for the homoskedastic DGP. Panel B displays analogous results, for size distortion  $s$ . The number of Monte Carlo simulations is 2000. The nominal coverage level is  $(1 - \alpha) = 0.90$ .

(independent of all the previous random variables both contemporaneously and at all leads and lags), and  $u_{1t} = au_{2t} + b_t$ , where  $a = \tilde{\alpha}/(1 + \theta^2)$ ,  $\tilde{\alpha} = -0.5$ , leading to  $E(u_{1t}u_{2t}) = \tilde{\alpha}$ . Define the  $(K_2 \times 1)$  coefficient vector  $\gamma$  as  $\gamma = (\gamma_1, 0, \dots, 0)'$ ,  $\gamma_1 = 0.5$ . Conditional heteroskedasticity is introduced by letting  $v_{1t} = Z_t' \gamma u_{1t}$  and  $v_{2t} = Z_t' \gamma u_{2t}$  be the  $t$ -th element of  $v_1$  and  $v_2$ , respectively.

In DGP 1, we set  $\theta = 0$  to make  $(Z'v_1, Z'v_2)'$  serially uncorrelated, while preserving conditional heteroskedasticity.

We performed Monte Carlo simulations for sample sizes  $T = \{100, 250, 500, 1000\}$ , with  $K_2 = \{1, 2, 3, 4\}$  instruments, and for various strengths of identification,  $N_{\text{TOLS}}(\beta, C, W) = \{0.05, 0.10, 0.3\}$  (in the case of  $K_2 = 4$  instruments, we specified  $N_{\text{TOLS}}(\beta, C, W) = \{-0.05, -0.10, -0.3\}$ , as the Nagar bias is non-positive in this case). We set  $\beta = 1$  in all cases. Furthermore, we specified  $C = (c_*^2, c_*, \dots, c_*)$ , and using Matlab's `fzero` or `fsolve` solver we determined the value of  $c_*$  such that (given  $\beta$  and  $W$ ) it implies the desired amount of Nagar bias. The specific  $C$  vectors, along with the derivation of the covariance matrix  $W$  can be found in Section C of the Online Appendix. The numerical optimization to calculate the bounds of the Nagar bias in the simulations (and later in the empirical examples) was performed using the augmented Lagrangian algorithm (Birgin and Martínez, 2008) with the PRAXIS subalgorithm (Brent, 1972) in the NLOpt package (Johnson, 2014) through the OPTI Toolbox interface (Currie and Wilson, 2012).

Results are reported in [Table 5](#), showing that our proposed confidence interval delivers valid (although conservative) coverage rates for both DGPs across different values of Nagar bias  $N_{\text{TSLs}}(\beta, C, W)$ , sample sizes  $T$ , and number of instruments  $K_2$ .

Table 5: Coverage rates for the Nagar bias

<b>Panel A. Heteroskedastic IV model</b>												
$T \setminus N_{\text{TSLs}}$	$K_2 = 1$			$K_2 = 2$			$K_2 = 3$			$K_2 = 4$		
	0.3	0.1	0.05	0.3	0.1	0.05	0.3	0.1	0.05	-0.3	-0.1	-0.05
100	0.94	0.90	0.86	0.92	0.90	0.89	0.94	0.95	0.96	0.93	0.90	0.80
250	0.96	0.94	0.93	0.96	0.95	0.95	0.96	0.96	0.97	0.97	0.97	0.95
500	0.96	0.96	0.95	0.96	0.96	0.96	0.97	0.98	0.98	0.98	0.98	0.97
1000	0.96	0.96	0.96	0.98	0.98	0.97	0.98	0.99	0.99	0.98	0.98	0.98
<b>Panel B. Heteroskedastic and autocorrelated IV model</b>												
$T \setminus N_{\text{TSLs}}$	$K_2 = 1$			$K_2 = 2$			$K_2 = 3$			$K_2 = 4$		
	0.3	0.1	0.05	0.3	0.1	0.05	0.3	0.1	0.05	-0.3	-0.1	-0.05
100	0.89	0.83	0.76	0.88	0.86	0.84	0.86	0.89	0.89	0.84	0.75	0.87
250	0.94	0.92	0.89	0.93	0.93	0.92	0.94	0.95	0.95	0.93	0.91	0.95
500	0.96	0.95	0.93	0.96	0.95	0.94	0.96	0.97	0.97	0.97	0.97	0.97
1000	0.97	0.96	0.96	0.97	0.97	0.97	0.97	0.98	0.98	0.97	0.98	0.98

*Note:* The upper panel shows the empirical coverage rates of the proposed confidence interval for the Nagar bias  $N_{\text{TSLs}}(\beta, C, W)$  for different sample sizes  $T$ , values of the Nagar bias, and number of instruments  $K_2$  for DGP 1 in [Section 4.2](#). The lower panel displays analogous results, based on DGP 2. The number of Monte Carlo simulations is 2000. The nominal coverage level is  $(1 - \alpha) = 0.90$ .

## 5 EMPIRICAL ANALYSIS

### 5.1 Estimating the Intertemporal Elasticity of Substitution

The intertemporal elasticity of substitution (IES) is often estimated using a linearized Euler equation, which is commonly derived as an optimality condition of the household's problem in modern macroeconomic models. We illustrate our proposed methodology by using the same specifications of the consumption Euler equation as Yogo (2004) and Montiel Olea and Pflueger (2013). The model is a linear IV model, and we consider both the homoskedastic, and the heteroskedastic and serially correlated cases.

In particular, the structural equation is either of the following:

$$\Delta c_{t+1} = v + \psi r_{t+1} + u_{t+1}, \quad (55)$$

$$r_{t+1} = \zeta + \psi^{-1} \Delta c_{t+1} + \eta_{t+1}, \quad (56)$$

where  $\Delta$  denotes the first difference operator,  $c_t$  is the logarithm of the level of

consumption,  $\Delta c_{t+1}$  is consumption growth, and  $r_{t+1}$  is a real asset return,  $\psi$  is the IES parameter,  $\nu$  and  $\xi$  are constants, while  $u_{t+1}$  and  $\eta_{t+1}$  are stochastic disturbances, which can be conditionally heteroskedastic or autocorrelated. Note that [eq. \(56\)](#) expresses the same relationship between consumption growth and returns as [eq. \(55\)](#), but often the estimates of  $\psi$  are vastly different between these two specifications. Yogo (2004) argued that weak identification can explain these contradicting results.

To facilitate the comparison between their results and ours, we borrow the quarterly data set used by Yogo (2004) and Montiel Olea and Pflueger (2013) focusing on the US between 1947:Q3 and 1998:Q4. In [eqs. \(55\)](#) and [\(56\)](#), we use real per capita consumption growth for  $\Delta c_{t+1}$ , and the real return on the short-term interest rate for  $r_{t+1}$ . As Yogo (2004) notes, by using instruments dated  $t - 1$ ,  $\psi$  or its reciprocal  $\psi^{-1}$  can be still identified even if asset returns or consumption are conditionally heteroskedastic. We use the same instruments as Montiel Olea and Pflueger (2013), notably consumption growth, nominal interest rate, inflation rate, and the logarithm of the dividend-price ratio. Section B of the Online Appendix contains a detailed description of the data.

The estimation results are summarized in [Table 6](#). Panel A reports results for the heteroskedastic and serially correlated linear IV model, while Panel B focuses on the homoskedastic IV model. Note that, by comparing the results on the left and right panels, the point estimates suggest contradicting values for  $\psi$ , an empirical result also emphasized by Yogo (2004) and Montiel Olea and Pflueger (2013).

Furthermore, for the specification in [eq. \(56\)](#) both the Montiel Olea and Pflueger (2013) and the Stock and Yogo (2005) methods signal weak instruments, and our confidence interval for the Nagar bias agrees with them. However, the results are different when considering the specification in [eq. \(55\)](#): according to the Stock and Yogo (2005) test, the instruments are strong if one is willing to tolerate 10% bias or size distortion, while they are weak when applying the Montiel Olea and Pflueger (2013) test with  $\tau = 10\%$  maximum relative bias. This confirms the latter authors' finding, that the test developed for the homoskedastic case can be misleading in the presence of heteroskedasticity or autocorrelation. Surprisingly, our analysis reveals that the confidence interval for the Nagar bias is  $[-0.00, 0.02]$ , signaling almost no bias. What could explain these seemingly conflicting results? Recall that the Montiel Olea and Pflueger (2013) method tests the Nagar bias of the TSLS estimator *relative* to a benchmark, while our confidence interval is directly applicable without the need to specify a reference bias. Hence, if the benchmark bias (which is not consistently estimable) itself is small, then this could resolve the seemingly different results. The low IES estimate and the corresponding negligible Nagar bias are in line with the results of Havránek (2015), who finds in a large-scale meta analysis of the literature that after correcting for publication bias, IES estimates based on macroeconomic data are centered around zero.

Table 6: Intertemporal elasticity of substitution

<b>Panel A. Heteroskedastic/serially correlated IV model</b>	IES $\psi$	IES $\psi^{-1}$
TLSLS estimate (standard error)	0.06 (0.098)	0.68 (0.813)
$CI_{0.95}^{N_{\text{TSLs}}}$	[−0.00, 0.02]	[20.28, 12695.44]
$\hat{F}_{\text{eff}}$	8.14	2.65
Critical value ( $\tau = 0.1$ )	15.49	13.99
Critical value ( $\tau = 0.3$ )	<b>7.75</b>	7.04
<b>Panel B. Homoskedastic IV model</b>	IES $\psi$	IES $\psi^{-1}$
TLSLS estimate (standard error)	0.06 (0.086)	0.68 (0.474)
95% Confidence Interval for bias	[0.021, 0.058]	[0.069, 0.786]
95% Confidence Interval for size distortion	[0.034, 0.090]	[0.105, 0.816]
F-statistic	15.53	2.93
Critical value (5% bias)	16.85	16.85
Critical value (10% bias)	<b>10.27</b>	10.27
Critical value (5% size distortion)	24.58	24.58
Critical value (10% size distortion)	<b>13.96</b>	13.96

*Note:* The table displays the estimation results of the consumption Euler equations with  $\Delta c_{t+1}$  regressed on  $r_{t+1}$  (specification IES  $\psi$  in eq. (55)), and  $r_{t+1}$  regressed on  $\Delta c_{t+1}$  (specification IES  $\psi^{-1}$  in eq. (56)). Panel A shows results based on the heteroskedastic and autocorrelated IV model: TLSLS point estimates and HAC standard errors (Newey and West’s (1987) HAC estimator with 6 lags, as in Montiel Olea and Pflueger (2013)), the 95% level confidence interval for the Nagar bias, along with the effective  $F$ -statistics  $\hat{F}_{\text{eff}}$  and the corresponding 5% critical values, allowing for  $\tau$  relative bias. The asymptotic covariance matrix  $W$  was estimated by the Newey and West (1987) HAC estimator, with 6 lags, as in Montiel Olea and Pflueger (2013). Panel B displays results based on the homoskedastic IV model: the 95% level confidence interval (based on the non-central  $\chi^2$  method) for the relative bias and size distortion (assuming a nominal 5% level Wald test), the Stock and Yogo (2005)  $F$ -statistics and the corresponding critical values (at the 5% significance level). In both panels, critical values in bold correspond to strong instruments according to the specific threshold.

## 5.2 Estimating fiscal multipliers by Local Projections–IV

As the second empirical example, we provide confidence intervals for the Nagar bias in a local projections–IV model. In their recent study, Ramey and Zubairy (2018) estimated both state-dependent and state-independent government spending multipliers for the US, using quarterly data in a sample period spanning 1889 – 2015. In this paper, we build on their analysis and estimate cumulative fiscal multipliers when the state of the economy corresponds to zero lower bound (ZLB) or non-ZLB (“normal”) periods, in addition to state-independent (“linear”) multipliers.

Ramey and Zubairy (2018) estimate the following structural equations for the ZLB,

non-ZLB and state-independent specifications:

$$\sum_{j=0}^h y_{t+j} = c_h^{\text{ZLB}} + \gamma_h^{\text{ZLB}} I_{t-1} + \phi_h^{\text{ZLB}}(L) z_{t-1} I_{t-1} + m_h^{\text{ZLB}} \sum_{j=0}^h g_{t+j} I_{t-1} + \omega_{t+h}^{\text{ZLB}}, \quad (57)$$

$$\begin{aligned} \sum_{j=0}^h y_{t+j} = & c_h^{\text{normal}} + \gamma_h^{\text{normal}} (1 - I_{t-1}) + \phi_h^{\text{normal}}(L) z_{t-1} (1 - I_{t-1}) \\ & + m_h^{\text{normal}} \sum_{j=0}^h g_{t+j} (1 - I_{t-1}) + \omega_{t+h}^{\text{normal}}, \end{aligned} \quad (58)$$

$$\sum_{j=0}^h y_{t+j} = c_h^{\text{linear}} + \phi_h^{\text{linear}}(L) z_{t-1} + m_h^{\text{linear}} \sum_{j=0}^h g_{t+j} + \omega_{t+h}^{\text{linear}}, \quad (59)$$

where  $\sum_{j=0}^h y_{t+j}$  is the sum of real GDP divided by potential GDP over periods  $t$  to  $t+h$ ;  $I_{t-1}$  is a dummy variable indicating the state of the economy when the shock hits ( $I_{t-1} = 1$  in the ZLB period and  $I_{t-1} = 0$  in the normal period);  $\sum_{j=0}^h g_{t+j}$  is the sum of real government spending divided by potential GDP between  $t$  and  $t+h$ ;  $z_{t-1}$  is the same vector of control variables as used by the original authors containing: real GDP over its potential level, real government spending over potential real GDP, and the defense news shock variable (introduced later) when it is used as an instrument. For  $s = \{\text{ZLB, normal, linear}\}$ ,  $c_h^s$ ,  $\gamma_h^s$  are scalar coefficients;  $\phi_h^s(L)$  are polynomials in the lag operator  $L$  up to four lags;  $m_h^s$  are the government spending multipliers, which are the structural parameters of interest. The error terms  $\omega_{t+h}^s$  are potentially serially correlated and heteroskedastic. For a detailed description of the data, we refer to Section B of the Online Appendix.

Ramey and Zubairy (2018) estimate the government spending multipliers at the 2 and 4 year horizons (corresponding to  $h = 7$  and  $h = 15$ , denoted by 2Y and 4Y) by LP-IV, instrumenting the cumulative government spending variable. As instruments, they use either the Blanchard and Perotti (2002) shock (current normalized government spending, denoted by "BP"), or Ramey's (2011) defense news shock series (rescaled by lagged GDP deflator times trend GDP, denoted by "News"), or both.

Table 7 reports the empirical results. The columns labeled "Estimates" are the same as in Ramey and Zubairy (2018) and display the TSLS estimates, together with their HAC standard errors in parentheses. We show estimates for both the state-dependent multipliers (the specifications labeled "ZLB" and "Normal", referring to the ZLB and non-ZLB periods, respectively) and the state-independent (labeled as "Linear") specifications. The columns labeled "Confidence intervals for bias" report the confidence intervals for the Nagar bias of the TSLS estimator. The last three columns contain of the Montiel Olea and Pflueger (2013) test statistics (cases of strong instruments in bold), along with the 5% level critical values corresponding to  $\tau = 0.1$

Table 7: Government spending multipliers

IV(s)	Horizon	Estimates			Confidence intervals for bias			$\widehat{F}_{\text{eff}}$ and c.v. ( $\tau = 0.1$ )		
		Linear	ZLB	Normal	Linear	ZLB	Normal	Linear	ZLB	Normal
News	2Y	0.66 (0.07)	0.77 (0.11)	0.63 (0.15)	[0.00, 0.13]	[0.00, 0.12]	[0.00, 0.04]	19.95 [23.11]	22.61 [23.11]	<b>43.68</b> [23.11]
	4Y	0.71 (0.04)	0.77 (0.06)	0.77 (0.38)	[0.00, 0.22]	[0.00, 0.54]	[-0.04, 0.20]	11.55 [23.11]	10.21 [23.11]	<b>24.06</b> [23.11]
BP	2Y	0.38 (0.11)	0.64 (0.03)	0.10 (0.11)	[-0.07, -0.01]	[-0.00, 0.00]	[-0.01, 0.00]	<b>36.72</b> [23.11]	<b>53.98</b> [23.11]	<b>70.60</b> [23.11]
	4Y	0.47 (0.11)	0.71 (0.03)	0.12 (0.12)	[-0.21, -0.00]	[-0.00, 0.02]	[-0.02, 0.01]	20.11 [23.11]	21.03 [23.11]	<b>36.44</b> [23.11]
News & BP	2Y	0.42 (0.10)	0.67 (0.03)	0.26 (0.10)	$\emptyset$	[-0.00, 0.00]	[-0.00, -0.00]	<b>37.85</b> [13.19]	<b>37.20</b> [13.56]	<b>37.99</b> [13.06]
	4Y	0.56 (0.08)	0.76 (0.04)	0.21 (0.14)	$\emptyset$	[-0.01, 0.00]	[-0.07, 0.00]	14.90 [15.46]	12.11 [15.89]	<b>19.43</b> [18.20]

*Note:* The columns labeled "Estimates" report TOLS point estimates of fiscal multipliers, and Newey–West (1987) standard errors in parentheses, with Newey and West’s (1994) automatic bandwidth selection. The columns labeled "Confidence intervals for bias" report the 95% confidence intervals for the Nagar bias. The last three columns report the effective  $F$ -statistics, and the 5% critical values in brackets corresponding to a maximum relative bias of  $\tau = 0.1$ . Significant effective  $F$ -statistics indicating strong instruments are in bold. Blocks labeled "News" ("BP") refer to using Ramey’s (Blanchard and Perotti’s) shock as instrument, while News & BP means using both instruments at the same time. 2Y and 4Y correspond to the 2-year-horizon and 4-year-horizon, respectively. The symbol  $\emptyset$  means an empty confidence interval for the Nagar bias.

maximum relative bias in parentheses.

Researchers might want to be informed of the true instrument strength in addition to the testing procedure when using the news shocks in the linear and ZLB specifications at the 2-year-horizon, or the Blanchard–Perotti shock in the same specifications at the 4-year-horizon. Given that in these cases the effective  $F$ -statistics are slightly below their critical values, the instruments are potentially weak, leading to biased point estimates.

Overall, we find negligible biases when estimating the state-dependent model, either using the Blanchard–Perotti or both the Blanchard–Perotti and the news shocks as instruments, while we find some positive bias when using only the news shock instrument. After correcting the TOLS point estimate by the confidence interval for the Nagar bias, the estimates in the zero lower bound period based on using only the news shock instrument are very similar to those obtained using the other instruments: they range between 0.65 and 0.77 for the two-year multiplier, and between 0.23 and 0.77 for the four-year one. These results demonstrate that our proposed confidence can indeed provide additional and useful information to researchers.

Turning to the linear, state-independent specification of the model in eq. (59), labeled "Linear" in the table, when using one instrument at a time, our confidence intervals imply some positive bias, especially at the 4-year-horizon when using the news shock instrument, and negative bias when using the Blanchard–Perotti instru-

ment. However, when using both the Blanchard–Perotti and the news shock series at the same time, our results point in the direction of the invalidity of the instruments, as the confidence set  $CI_{1-\alpha}^{C,\beta}$  is empty, meaning there is no  $(\tilde{C}, \tilde{\beta}) \in \mathbb{R}^3$  which would be consistent with the model. This was also mentioned by Ramey and Zubairy (2018), who note in their Footnote 36 that the overidentifying restrictions are rejected.

## 6 CONCLUSION

In this paper we propose confidence intervals for the strength of identification, and in particular, bias and size distortion in the homoskedastic IV model, and Nagar bias in the heteroskedastic/autocorrelated linear IV model as well as local projections–IV models. Our proposed methodologies allow researchers working with either microeconomic or macroeconomic data to determine how strong their instruments are and how big their size distortion and bias can be. The practical implementation of our proposed methodologies has the benefit of being easy and computationally simple. Monte Carlo simulations show that the proposed confidence intervals have correct coverage even for moderate sample sizes.

The application of our new methodology uncovers a series of interesting empirical facts. In particular, our analysis of the consumption Euler equation confirms that weak identification poses a serious challenge to estimating the intertemporal elasticity of substitution parameter. However, in one model specification, our results suggest that the bias of the point estimate might be minor, and the available testing procedure only implies weak instruments due to its formulation in terms of a benchmark bias. In contrast, our method is applicable without reference to such a benchmark bias. Furthermore, our local projections–IV analysis shows that the presence of biases can help reconcile the differences in the fiscal policy multipliers across different sets of instruments in the zero lower bound period.

## SUPPLEMENTARY MATERIALS

The online supplementary materials contain the proofs, further theoretical and Monte Carlo results, a description of the data sets used in the present paper, and all the code to replicate our results.

## REFERENCES

- Andrews, D. W. K., Moreira, M. J., and Stock J. H. (2006), "Optimal Two-sided Invariant Similar Tests for Instrumental Variables Regression," *Econometrica*, 74, 715-752.
- Angrist, J., and Krueger A. (1991), "Does Compulsory School Attendance Affect Schooling and Earnings?," *Quarterly Journal of Economics*, 106, 979-1014.
- Birgin, E. G., and Martínez, J. M. (2008), "Improving ultimate convergence of an augmented Lagrangian method," *Optimization Methods and Software*, 23, 177-195.
- Blanchard, O., and Perotti, R. (2002), "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output," *Quarterly Journal of Economics*, 117, 1329-1368.
- Bound, J., Jaeger, D., and Baker, R. (1995), "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak," *Journal of the American Statistical Association*, 90, 443-450.
- Brent, R. (1972), "Algorithms for Minimization without Derivatives," Prentice Hall (reprinted by Dover, 2002).
- Cragg, J. G., and Donald, S. G. (1993), "Testing identifiability and specification in instrumental variable models," *Econometric Theory*, 9, 222-240.
- Dufour, J.-M. (1997), "Some Impossibility Theorems in Econometrics With Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365-1387.
- Hansen, B. E. (1999), "The grid bootstrap and the autoregressive model. *The Review of Economics and Statistics*, 81, 594-607.
- Havránek, T. (2015), "Measuring Intertemporal Substitution: The Importance of Method Choices and Selective Reporting," *Journal of the European Economic Association*, 13, 1180-1204.
- Johnson, S. G., (2014), "The NLOpt nonlinear-optimization package", URL: <http://ab-initio.mit.edu/nlopt>.
- Jordà, Ò. (2005), "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, 95, 161-182.
- Jordà, Ò., Schularick, M., and Taylor, A. M. (2015), "Betting the house," *Journal of International Economics*, 96, S2-S18.
- Kent, J. T., and Hainsworth, T. J (1995), "Confidence intervals for the noncentral chi-squared distribution," *Journal of Statistical Planning and Inference*, 46, 147-159.
- Currie, J., and Wilson, D. I. (2012), "OPTI: Lowering the Barrier Between Open Source Optimizers and the Industrial MATLAB User," in *Foundations of Computer-Aided Process Operations*, ed. Sahinidis, N., and Pinto, J., Savannah, Georgia, USA.
- Montiel Olea, J. L., and Pflueger, C. (2013), "A robust test for weak instruments,"

*Journal of Business & Economic Statistics*, 31, 358-369.

Nagar, A. L. (1959), "The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations," *Econometrica*, 27, 575-595.

Newey, W. K., and West, K. D. (1987), "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, 55, 703-708.

Newey, W. K., and West, K. D. (1994), "Automatic Lag Selection in Covariance Matrix Estimation," *The Review of Economic Studies*, 55, 631-653.

Ramey, V. A. (2011), "Identifying Government Spending Shocks: It's All in the Timing," *Quarterly Journal of Economics*, 126, 51-102.

Ramey, V. A., and Zubairy, S. (2018), "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data," *Journal of Political Economy*, 126, 850-901.

Skeels, C. L., and Windmeijer, F. (2016), "On the Stock-Yogo Tables," Bristol Economics Discussion Papers 16/679, Department of Economics, University of Bristol.

Staiger, D., and Stock, J. H. (1997), "Instrumental variables regression with weak instruments," *Econometrica*, 65, 557-586.

Stock, J. H., Wright, J. H., and Yogo, M. (2002), "A survey of weak instruments and weak identification in generalized method of moments," *Journal of Business & Economic Statistics*, 20, 518-529.

Stock, J. H., and Yogo, M. (2005), "Testing for weak instruments in linear IV regression," in *Identification and Inference for Econometric Models*, ed. Andrews, D. W. K., Cambridge University Press, New York, pp. 80-108.

Stock, J. H., and Watson, M. W. (2018), "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments," *The Economic Journal*, 128, 917-948.

Wasserstein, R. L., and Lazar, N. A. (2016), "The ASA's Statement on p-Values: Context, Process, and Purpose," *The American Statistician*, 70, 129-133.

Yogo, M. (2004), "Estimating the elasticity of intertemporal substitution when instruments are weak," *The Review of Economics and Statistics*, 86, 797-810.