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

Gergely Ganics[1], Atsushi Inoue[2] and Barbara Rossi[3]

August 4, 2019

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 variables models with heteroskedasticity and serial correlation. 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 consistently estimated, 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, albeit conservative, in some cases, small sample coverage. We illustrate the usefulness of the proposed methods 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.

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

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

[3]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.

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, allowing for heteroskedasticity and serial correlation when the model contains one endogenous regressor. 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 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 sufficiently large F-statistic 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 and size distortion 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, as 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: "(...) report standard errors in parentheses but do not use *s to report significance levels.")

In our framework, 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 into the appropriate expression for the strength of identification. To construct confidence intervals for the bias and the size distortion, 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 and the Wald test's size distortion directly through the projection method. We note that in general, the projection method leads to conservative confidence intervals, thus we recommend the non-central χ^2 method over the projection method in the homoskedastic case with one endogenous regressor.

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 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, using a Heteroskedasticity and Autocorrelation Consistent (HAC) estimator. Monte Carlo simulations demonstrate that our methods have good, albeit conservative, in some cases, coverage.

We illustrate the usefulness of our methodology in two empirical applications. In the first one, we the 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. Andrews, Stock and Sun (in press) provide an excellent survey of the recent literature. 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 (the computational problems increase when the number of endogenous variables increases) and is only available in special cases. We note that the Anderson-Rubin statistic has a straightforward implementation in just-identified models with one endogenous regressor (in both homoskedastic and heteroskedastic/autocorrelated cases), and even in set-ups where the parameters of interest are nonlinear as in the weak structural vector autoregressive – IV (SVAR-IV) model of Montiel Olea, Stock, Watson (2018). Tests for weak instruments can be computationally less 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. Section 3 describes our proposed confidence intervals. Section 4 provides Monte Carlo simulation results. Section 5 presents empirical results, and Section 6 concludes. Throughout the paper, T denotes the sample size, $\underset{p}{\rightarrow}$ and $\underset{d}{\rightarrow}$ stand for convergence in probability and in distribution, respectively.

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, Y and Z are $(T \times 1)$ vectors of the endogenous regressor and instrument; u and V are $(T \times 1)$ vectors of independent, mean-zero disturbances, the latter with variance σ_{VV}. For simplicity, σ_{VV} and $\mathrm{E}(Z_t^2)$ are known. The structural equation is eq. (1), with the structural coefficient of interest β. Information on the strength of the instrument is carried by the parameter Π in eq. (2).

2.1 Confidence Intervals for the Strength of Identification

In the case of one endogenous variable, 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, the test statistic's critical values are different from standard values based on the chi-squared distribution, thus making inference difficult.

Let $\widehat{\Pi}_T = (Z'Z)^{-1}(Z'Y)$ denote the OLS estimator of Π 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):

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

$$\sqrt{T}\left(\widehat{\Pi}_T - 0\right) = \sqrt{T}\,(Z'Z)^{-1}\,(Z'Y) = \left(\frac{Z'Z}{T}\right)^{-1}\left(\frac{Z'Z}{T}C\right) + \left(\frac{Z'Z}{T}\right)^{-1}\left(\frac{Z'V}{\sqrt{T}}\right) \underset{d}{\rightarrow} C + \nu\,, \tag{4}$$

where $\nu = \mathrm{E}\left(Z_t^2\right)^{-1}\Psi_{ZV}$, Ψ_{ZV} is a random variable whose distribution is $\mathcal{N}\left(0, \mathrm{E}\left(Z_t^2\right)\sigma_{VV}\right)$. 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:

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

$$= \left(\frac{Z'Z}{T}\right)^{-1}\left(\frac{Z'Z}{T}C\right) + \left(\frac{Z'Z}{T}\right)^{-1}\left(\frac{Z'V}{\sqrt{T}}\right) - C \underset{d}{\rightarrow} \nu\,. \tag{6}$$

This result implies that

$$\mathcal{F}_{\Pi} \equiv \frac{\left(\widehat{\Pi}_T - \Pi\right)^2}{\sigma_{VV}\left(Z'Z\right)^{-1}} \underset{d}{\rightarrow} \chi^2_1\,, \tag{7}$$

where χ^2_1 denotes a chi-squared distribution with one degree of freedom. Thus, one conveniently obtains a confidence interval for C by inverting a standard χ^2_1 distribution.

It might be surprising that the confidence intervals that we propose can be obtained by inverting limiting standard chi-squared distributions while the usual test statistic $\mathcal{F}_0$ cannot be used for this purpose. The intuition is that the first-stage F-statistic 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 where the parameter is local to the null hypothesis, such as confidence intervals for highly persistent (local-to-unity) autoregressive processes.

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 TSLS estimator of β is a function of the concentration parameter:

$$\mu_1^2 = C^2\,\mathrm{E}(Z_t^2)/\sigma_{VV}\,, \tag{8}$$

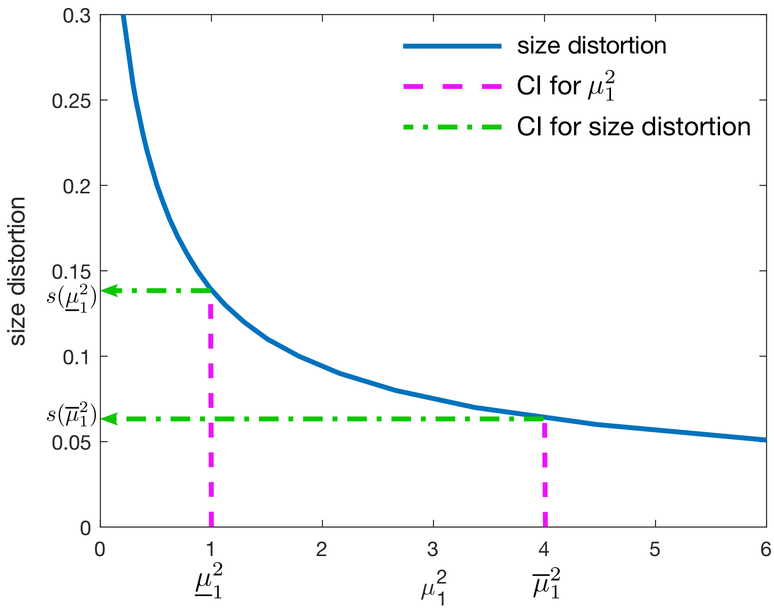
where the subscript "1" refers to the number of instruments. Let us define $s\left(\mu_1^2\right)$ to be the size distortion. Figure 1 shows the size distortion as a function of μ_1^2 (the nominal level of the Wald test is 5%).

Note that once one has a confidence interval for C, CI_C, one directly obtains a confidence interval for μ_1^2, $\mathrm{CI}_{\mu_1^2}$, as follows:

$$\mathrm{CI}_{\mu_1^2} = \left\{\widetilde{\mu}_1^2 = \widetilde{C}^2\,\mathrm{E}(Z_t^2)/\sigma_{VV} \text{ such that } \widetilde{C} \in \mathrm{CI}_C\right\}. \tag{9}$$

One can then construct a confidence interval for $s\left(\mu_1^2\right)$ by a non-central χ^2 method or the projection method. In the latter case, suppose $\mathrm{CI}_C = \left[\underline{C}, \overline{C}\right]$ is the confidence interval for C obtained by inverting the χ_1^2 distribution, where $\underline{C} > 0$. Then $\mathrm{CI}_{\mu_1^2} = \left[\underline{\mu}_1^2 = \underline{C}^2\,\mathrm{E}(Z_t^2)/\sigma_{VV}\,, \overline{\mu}_1^2 = \overline{C}^2\,\mathrm{E}(Z_t^2)/\sigma_{VV}\right]$. Suppose that $\mathrm{CI}_{\mu_1^2} = [1, 4]$ is the confidence interval for μ_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 $[0.06, 0.14]$.

Figure 1: Construction of confidence interval for size distortion



Note: The figure plots size distortion as a function of μ_1^2 (solid line). The confidence interval for μ_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 ECONOMETRIC FRAMEWORKS

In this section, we describe the econometric frameworks we consider, and the confidence intervals that we propose. The Euclidean norm of a vector a is denoted by $\|a\|$, $\mathrm{tr}\,(\cdot)$ is the trace operator, $\mathrm{vec}\,(\cdot)$ 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 ψ and covariance matrix Ξ, and χ^2_k denotes the chi-squared distribution with k degrees of freedom. For any $(T \times K)$ matrix A, $P_A \equiv A\,(A'A)^{-1}\,A'$, and $M_A = I_T - P_A$, where I_T is the $(T \times T)$ identity matrix. 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.

3.1 The General Linear IV Model

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

$$y = Y\beta + X\gamma + u, \tag{10}$$

$$Y = Z\Pi + X\Phi + V, \tag{11}$$

where y is a $(T \times 1)$ vector and Y is a $(T \times n)$ matrix of included endogenous regressors. 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). β is an $(n \times 1)$, while γ is a $(K_1 \times 1)$ vector of coefficients. Π is a matrix of coefficients of dimension $(K_2 \times n)$, and Φ 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. The precise assumptions are stated later.

We define $X_t = (X_{1t}, \ldots, X_{K_1 t})'$, $Z_t = (Z_{1t}, \ldots, Z_{K_2 t})'$, $V_t = (V_{1t}, \ldots, V_{nt})'$, $\underline{Z}_t = (X_t', Z_t')'$ as the vectors of the t-th observations of the respective variables, $t = 1, \ldots, T$, and $\underline{Z} = [X\,Z]$. In order to develop our asymptotic theory, it is convenient to project out the exogenous regressors, X. That is, let $Y^\perp \equiv M_X Y$, $Z^\perp \equiv M_X Z$, $V^\perp \equiv M_X V$, and $u^\perp \equiv M_X u$. Moreover, let $V_t^\perp$ be the transpose of the t-th row of $V^\perp$, and similarly for $Z_t^\perp$ and $u_t^\perp$. Using this notation, we can rewrite eqs. (10) and (11) as:

$$y^\perp = Y^\perp\beta + u^\perp\,, \tag{12}$$

$$Y^\perp = Z^\perp\Pi + V^\perp\,. \tag{13}$$

3.2 The Linear IV Model With $n = 1$ Endogenous Regressor

In empirical applications, the linear IV model with $n = 1$ endogenous regressor is of particular interest. This subsection first introduces our proposed confidence intervals in the homoskedastic setting, and then the heteroskedastic/autocorrelated set-up.

3.2.1 Homoskedastic IV model

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. 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.

In this section, we keep the generality of the matrix notation when possible, to avoid having re-introducing it for the general case of $n \geq 1$ endogenous variables later.

Let us define the population second moment matrices Σ and Q are as follows:

$$\Sigma = E\left[\begin{pmatrix} u_t \\ V_t \end{pmatrix}\begin{pmatrix} u_t & V_t' \end{pmatrix}\right] = \begin{bmatrix} \sigma_{uu} & \Sigma_{uV} \\ \Sigma_{Vu} & \Sigma_{VV} \end{bmatrix}, \tag{14}$$

$$Q = E\left(\underline{Z}_t \underline{Z}_t'\right) = \begin{bmatrix} Q_{XX} & Q_{XZ} \\ Q_{ZX} & Q_{ZZ} \end{bmatrix}. \tag{15}$$

In this section we make the same assumptions as SSY.

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

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

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

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

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

Assumption L_Π models Π 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 3.2.2.

Furthermore, let us define $\Omega \equiv Q_{ZZ} - Q_{ZX}Q_{XX}^{-1}Q_{XZ} = Q_{Z^\perp Z^\perp}$, where $Q_{Z^\perp Z^\perp} \equiv E(Z_t^\perp Z_t^{\perp\prime})$, and $\widehat{\Omega} \equiv Z^{\perp\prime}Z^\perp/T$. Moreover, let $\widehat{\Pi}_T \equiv (Z^{\perp\prime}Z^\perp)^{-1}Z^{\perp\prime}Y^\perp$ denote the OLS estimator of Π in eq. (13). Note that, by the exogeneity of X, $E\left(X_t V_t'\right) = 0$, thus $\Sigma_{V^\perp V^\perp} \equiv E\left(V_t^\perp V_t^{\perp\prime}\right) = \Sigma_{VV}$, where $\Sigma_{VV} \equiv E(V_t V_t')$. Additionally, Assumption M implies $\widehat{\Omega} \underset{p}{\rightarrow} \Omega$, and that $\widehat{\Sigma}_{VV} \equiv Y^{\perp\prime} M_{Z^{\perp}} Y^{\perp} / (T - K_1 - K_2) \underset{p}{\rightarrow} \Sigma_{VV}$.

The concentration parameter plays an important role in the construction of the confidence intervals for the strength of identification. In the case of $n = 1$ endogenous regressor, Σ_{VV} is a scalar σ_{VV} (whose consistent estimator is the same, $\widehat{\sigma}_{VV} = \widehat{\Sigma}_{VV}$), and the scalar concentration parameter is given by:

$$\mu^2_{K_2} \equiv \frac{1}{K_2} C' \Omega C / \sigma_{VV} \,. \tag{16}$$

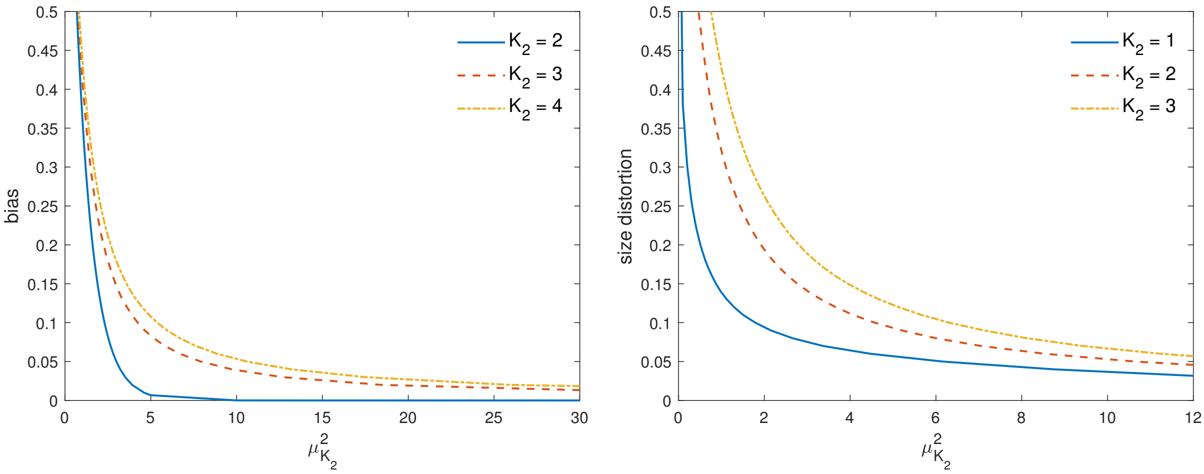
As Stock and Yogo (2005) demonstrated, the (*i*) worst-case asymptotic bias relative to the OLS estimator or (*ii*) worst-case asymptotic size distortion of the Wald test on β – where the worst-case corresponds to the maximum of these quantities over all possible degrees of simultaneity between the error 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 concentration parameter. In this subsection, 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. Finally, we note that the concentration parameter's multivariate generalization (in the case of $n \geq 1$ endogenous regressors) is the concentration matrix, as explained later.

Let $b(\mu^2_{K_2}; n, K_2)$ and $s(\mu^2_{K_2}; n, K_2)$ denote the bias and the size distortion of the TSLS estimator, respectively, as the function of $\mu^2_{K_2}$ 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 analogous results for $n = \{1, 2, 3\}$ endogenous variables and $K_2 = n + 1, \ldots, 30$ (bias) and $K_2 = n, \ldots, 30$ (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 β has a nominal level of 5%. Using the MATLAB code that we provide, the simulations can be performed at a variety of nominal levels.

The starting point of our proposed confidence interval is the asymptotic distribution of the OLS estimator of Π in eq. (13). Under Assumptions L_{Π} and M, the asymptotic distribution of $\widehat{\Pi}_T$ is given by

$$\sqrt{T} \widehat{\Pi}_T \underset{d}{\rightarrow} \mathcal{N}(C, \sigma_{VV} \Omega^{-1}) \,, \tag{17}$$

Figure 2: Bias and size distortion of TSLS estimator as a function of $\mu^2_{K_2}$ $(n = 1)$



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 appoach of Stock and Yogo (2005).

which by Slutsky's theorem implies that

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

$$f_T \equiv m_T' m_T \underset{d}{\to} \chi^2_{K_2}\left(K_2\mu^2_{K_2}\right); \tag{19}$$

that is, f_T asymptotically follows the non-central chi-squared distribution with K_2 degrees of freedom and non-centrality parameter $K_2\mu^2_{K_2}$. By obtaining a confidence set for $\mu^2_{K_2}$ 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 $\mu^2_{K_2}$, through $b(\mu^2_{K_2}; n, K_2)$ and $s(\mu^2_{K_2}; n, K_2)$, 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 $F_{K_2}(x, K_2\mu^2_{K_2})$ denote the cumulative distribution function (CDF) of the non-central chi-squared distribution with K_2 degrees of freedom and non-centrality parameter $K_2\mu^2_{K_2}$ evaluated at x, and let $F^{-1}_{K_2}(q, K_2\mu^2_{K_2})$ denote the corresponding quantile function evaluated at q. Then the following algorithm leads to $(1-\alpha)$ level asymptotic confidence intervals for $\mu^2_{K_2}$.

1. Lower bound: If $\sqrt{f_T} \leq \sqrt{F^{-1}_{K_2}(1-\alpha, 0)}$, then set $l^{\mu^2_{K_2}}_{1-\alpha} = 0$. 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 $l^{\mu^2_{K_2}}_{1-\alpha} = (\sqrt{f_T} - b^*)^2 / K_2$.
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 $\mathrm{CI}_{1-\alpha}^{\mu_{K_2}^2} \equiv [l_{1-\alpha}^{\mu_{K_2}^2}, u_{1-\alpha}^{\mu_{K_2}^2}]$ is a $(1-\alpha)$ level asymptotic confidence interval for $\mu_{K_2}^2$. Let us define

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

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

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 based on the non-central χ^2 distribution): Under Assumptions L_Π and M, $\mathrm{CI}_{1-\alpha}^{\mu_{K_2}^2}$ is an asymptotically valid $(1-\alpha)$ level confidence interval for $\mu_{K_2}^2$, that is,

$$\lim_{T\to\infty} P\left(\mu_{K_2}^2 \in \mathrm{CI}_{1-\alpha}^{\mu_{K_2}^2}\right) = 1-\alpha\,. \tag{22}$$

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:

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

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

Proof. See Section A of the Online Appendix. □

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).

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 β more biased than a pre-specified tolerance, or controlling that the Wald test on β does not display higher size distortion than a threshold. Their theory builds on the asymptotic distribution of the first-stage F-statistic (or its multivariate generalization when $n \geq 1$). 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 $\mu^2_{K_2}$, 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* $(Z_t, V_t)' \sim iid\ \mathcal{N}(0, I_2)$, $\Pi = \Pi_T = C/\sqrt{T}$, $C = 0.5$, $T = 100$, $X_t = 1$ *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 percentiles of the t-statistic obtained using Hansen's (1999) grid bootstrap are straight lines (i.e. independent of* Π*) and equal to* ± 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* Π_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. The results confirm that the simulated quantiles of the t-statistic are virtually indistinguishable from their asymptotic counterparts (*± 1.64*) and constant over* Π *(see Section G of the Online Appendix).*

Example 2. *To illustrate our methodology in an empirical setting, consider Angrist and Krueger's (1991) problem of estimating the returns to education 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 distortion. 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 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 that when performing a Wald test on* β *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 χ^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 χ^2 – based confidence intervals suggest a 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

	Angrist and Krueger (1991) ($K_2 = 28$)	*Bound et al. (1995)* ($K_2 = 3$)
TSLS 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 χ^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 χ^2	$[0.777, 0.950]$	$[0.030, 0.097]$
F-statistic	1.61	13.49
Critical value (5% bias)	21.42	13.91
Critical value (10% bias)	11.34	**9.08**
Critical value (5% size distortion)	81.40	22.30
Critical value (10% size distortion)	42.37	**12.83**

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.

3.2.2 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.

Following Montiel Olea and Pflueger (2013), consider the linear IV model:

$$y^{\perp} = Z^{\perp}\Pi\beta + v_1\,, \tag{25}$$

$$Y^{\perp} = Z^{\perp}\Pi + v_2\,, \tag{26}$$

where eq. (25) is the structural equation of interest in reduced form, while eq. (26) 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, β is a scalar coefficient, Π 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 $Z^{\perp\prime}Z^{\perp}/T = I_{K_2}$.

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

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

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

(b) $[v_1\, v_2]'[v_1\, v_2]/T \underset{p}{\to} \kappa$ *for some positive definite κ;*

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

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) focus on the Nagar (1959) bias, defined as

$$N_{\text{TSLS}}(\beta, C, W) \equiv \mu^{-2}\frac{\text{tr}(\omega_{12})}{\text{tr}(\omega_2)}\left[1 - 2\frac{C_0'\omega_{12}C_0}{\text{tr}(\omega_{12})}\right]\,, \tag{27}$$

where $C = \|C\|C_0$, $\mu^2 \equiv \|C\|^2/\operatorname{tr}(W_2)$, $\omega_1 \equiv W_1 - 2\beta W_{12} + \beta^2 W_2$, $\omega_{12} \equiv W_{12} - \beta W_2$, and $\omega_2 \equiv W_2$. Note that μ^2 can be thought of as the analog of the concentration parameter $\mu^2_{K_2}$ 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 $\text{BM}(\beta, W) \equiv \sqrt{\operatorname{tr}(\omega_1)/\operatorname{tr}(\omega_2)}$, 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

$$\mu^2 \in \mathbb{R}_+ : \sup_{\beta \in \mathbb{R}, C_0 \in \mathcal{S}^{K_2-1}} \frac{|N_{\text{TSLS}}(\beta, \mu\sqrt{\text{tr}(W_2)}C_0, W)|}{\text{BM}(\beta, W)} > \tau\,, \tag{28}$$

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 τ of the benchmark bias $\text{BM}(\beta, W)$ for at least some value of the structural parameter β and some direction of the first-stage coefficients C_0. Montiel Olea and Pflueger (2013) propose the so-called *effective first-stage F-statistic* to test the null hypothesis of weak instruments:

$$\widehat{F}_{\text{eff}} \equiv \frac{Y^{\perp\prime}(Z^{\perp}Z^{\perp\prime}/T)Y^{\perp}}{\text{tr}(\widehat{W}_2)}\,. \tag{29}$$

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. (27). 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, a further complication arises due to the fact that the Nagar bias depends on the structural parameter β, 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 by combining two ideas: the method of obtaining a joint confidence set for C and β, and using a consistent estimate of W. Consider a compact expression of eqs. (25) and (26) (e.g., Andrews, Moreira and Stock (2006)):

$$\widetilde{Y} = Z^{\perp}\Pi a' + \tilde{v}\,, \tag{30}$$

where $\widetilde{Y} \equiv [y^{\perp},\, Y^{\perp}]$, $a \equiv (\beta, 1)'$ and $\tilde{v} \equiv [v_1,\, v_2]$. Note that the coefficient matrix $\Pi a'$ has an interesting structure: its first column is $\Pi\beta$, while its second column is Π. Let us define its vectorized version as $\Gamma \equiv [\Pi'\beta,\, \Pi']'$, then vectorize eq. (30):

$$\text{vec}\left(\widetilde{Y}\right) = (I_2 \otimes Z^{\perp})\Gamma + \text{vec}(\tilde{v})\,. \tag{31}$$

Consider the asymptotic distribution of the OLS estimator of Γ in eq. (31):

$$\sqrt{T}\left(\widehat{\Gamma} - \Gamma\right) = \begin{bmatrix} T^{-1/2}Z^{\perp\prime}v_1 \\ T^{-1/2}Z^{\perp\prime}v_2 \end{bmatrix} = \begin{bmatrix} \widehat{\psi} - C\beta \\ \widehat{C} - C \end{bmatrix} \underset{d}{\to} \mathcal{N}(0, W)\,, \tag{32}$$

where $\widehat{\psi}$ is $\sqrt{T}$ times the OLS estimator of $\Pi_T\beta$ in the structural equation, $\widehat{C} = \widehat{\Pi}_T\sqrt{T}$, and we used the normalization $Z^{\perp\prime}Z^{\perp}/T = I_{K_2}$ and Assumptions HL and L_{Π}. 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

$$\mathcal{W}(C,\beta) \equiv \begin{bmatrix} \widehat{\psi} - C\beta \\ \widehat{C} - C \end{bmatrix}' \widehat{W}^{-1} \begin{bmatrix} \widehat{\psi} - C\beta \\ \widehat{C} - C \end{bmatrix} \underset{d}{\to} \chi^2_{2K_2}\,. \tag{33}$$

By taking the $(1-\alpha)$ quantile of the $\chi^2_{2K_2}$ distribution (denoted by $\chi^2_{2K_2,1-\alpha}$), the Wald statistic $\mathcal{W}(C,\beta)$ can be inverted to obtain an asymptotically valid $(1-\alpha)$ level *joint* confidence set for C and β, formally:

$$\mathrm{CI}^{C,\beta}_{1-\alpha} \equiv \left\{\forall\,(\widetilde{C},\widetilde{\beta}) \in \mathbb{R}^{K_2+1} : \mathcal{W}(\widetilde{C},\widetilde{\beta}) \le \chi^2_{2K_2,1-\alpha}\right\}\,. \tag{34}$$

Note that if $\widetilde{C} = 0$ is in the confidence set $\mathrm{CI}^{C,\beta}_{1-\alpha}$, then the confidence set for β is unbounded, which is in line with the findings of Dufour (1997). If this is the case, then the instruments are very weak indeed, and β might not be identified. Therefore when $\widetilde{C} = 0 \in \mathrm{CI}^{C,\beta}_{1-\alpha}$, we take $[-\infty,+\infty]$ as our confidence set for the Nagar bias. Another peculiar case is when the confidence set $\mathrm{CI}^{C,\beta}_{1-\alpha}$ is empty, which can happen when a confidence set is based on the inversion principle. This situation indicates that the data reject 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.

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

$$\widetilde{N}_{\mathrm{TSLS}}(\widetilde{\beta},\widetilde{C},\widehat{W}) \equiv \widetilde{\mu}^{-2}\frac{\mathrm{tr}(\widetilde{\omega}_{12})}{\mathrm{tr}(\widetilde{\omega}_2)}\left[1 - 2\frac{\widetilde{C}_0'\widetilde{\omega}_{12}\widetilde{C}_0}{\mathrm{tr}(\widetilde{\omega}_{12})}\right], \tag{35}$$

where $\widetilde{C} = \|\widetilde{C}\|\widetilde{C}_0$, $\widetilde{\mu}^2 \equiv \|\widetilde{C}\|^2/\,\mathrm{tr}(\widehat{W}_2)$, $\widetilde{\omega}_1 \equiv \widehat{W}_1 - 2\widetilde{\beta}\widehat{W}_{12} + \widetilde{\beta}^2\widehat{W}_2$, $\widetilde{\omega}_{12} \equiv \widehat{W}_{12} - \widetilde{\beta}\widehat{W}_2$, and $\widetilde{\omega}_2 \equiv \widehat{W}_2$.

Let us define $L^N_{1-\alpha} \equiv \min_{(\widetilde{C},\widetilde{\beta})\in\mathrm{CI}^{C,\beta}_{1-\alpha}} \widetilde{N}_{\mathrm{TSLS}}(\widetilde{\beta},\widetilde{C},\widehat{W})$ and $U^N_{1-\alpha} \equiv \max_{(\widetilde{C},\widetilde{\beta})\in\mathrm{CI}^{C,\beta}_{1-\alpha}} \widetilde{N}_{\mathrm{TSLS}}(\widetilde{\beta},\widetilde{C},\widehat{W})$. Our proposed $(1-\alpha)$ level asymptotic confidence interval for $N_{\mathrm{TSLS}}(\beta,C,W)$ is

$$\mathrm{CI}^{N_{\mathrm{TSLS}}}_{1-\alpha} = \begin{cases} \left[L^N_{1-\alpha}, U^N_{1-\alpha}\right] & \text{if } \mathrm{CI}^{C,\beta}_{1-\alpha} \neq \emptyset \text{ and } \widetilde{C} = 0 \notin \mathrm{CI}^{C,\beta}_{1-\alpha}\,, \\ [-\infty,+\infty] & \text{if } \widetilde{C} = 0 \in \mathrm{CI}^{C,\beta}_{1-\alpha}\,, \\ \emptyset & \text{if } \mathrm{CI}^{C,\beta}_{1-\alpha} = \emptyset\,. \end{cases} \tag{36}$$

We summarize our results in the following proposition.

Proposition 2 (Confidence interval for the Nagar bias): Under Assumptions L_{Π} and

HL, $\mathrm{CI}_{1-\alpha}^{N_{\mathrm{TSLS}}}$ in eq. (36) is an asymptotically valid confidence interval for the Nagar bias $N_{\mathrm{TSLS}}(\beta, C, W)$, that is

$$\lim_{T\to\infty} \mathrm{P}\left(N_{\mathrm{TSLS}}(\beta, C, W) \in \mathrm{CI}_{1-\alpha}^{N_{\mathrm{TSLS}}}\right) \geq 1-\alpha\,. \tag{37}$$

Proof. See Section A of the Online Appendix. □

Note that, as $\widetilde{N}_{\mathrm{TSLS}}(\widetilde{\beta}, \widetilde{C}, \widehat{W})$ is not a one-to-one function of $\left(\widetilde{C}, \widetilde{\beta}\right)$ in general, our proposed confidence interval may be conservative.

In addition to learning about the bias of the TSLS estimator in the (potentially) heteroskedastic/autocorrelated IV model, researchers might want to perform a test on the structural parameter β. In what follows, we provide a confidence interval for the size distortion of the Wald test on β. Compared to the *worst-case* size distortion investigated by Stock and Yogo (2005), we provide a confidence intervals for the *actual* asymptotic size distortion of the Wald test, and allow for heteroskedasticity and autocorrelation, although we consider the case of $n = 1$ endogenous regressor only.

Consider the TSLS estimator of β, given by

$$\widehat{\beta} = \left(Y^{\perp\prime} Z^{\perp} Z^{\perp\prime} Y^{\perp}\right)^{-1} \left(Y^{\perp\prime} Z^{\perp} Z^{\perp\prime} y^{\perp\prime}\right)\,, \tag{38}$$

from which it follows that

$$\sqrt{T}\left(\widehat{\beta}-\beta\right) = \left(\frac{Y^{\perp\prime}Z^{\perp}}{T}\frac{Z^{\perp\prime}Y^{\perp}}{T}\right)^{-1}\left(\frac{Y^{\perp\prime}Z^{\perp}}{T}\frac{Z^{\perp\prime}u^{\perp}}{\sqrt{T}}\right)\,. \tag{39}$$

Let S denote the long-run variance of $Z_t^{\perp} u_t^{\perp}$, that is $S \equiv \lim_{T\to\infty} \mathrm{Var}\left(T^{-1/2} Z^{\perp\prime} u^{\perp}\right)$, and $\widehat{S}$ an estimator of S (to be specified later).

The Wald statistic testing a null hypothesis on β is

$$\mathbb{W}_T \equiv \frac{T\left(\widehat{\beta}-\beta\right)^2}{\left(\frac{Y^{\perp\prime}Z^{\perp}}{T}\frac{Z^{\perp\prime}Y^{\perp}}{T}\right)^{-1}\frac{Y^{\perp\prime}Z^{\perp}}{T}\widehat{S}\frac{Z^{\perp\prime}Y^{\perp}}{T}\left(\frac{Y^{\perp\prime}Z^{\perp}}{T}\frac{Z^{\perp\prime}Y^{\perp}}{T}\right)^{-1}} = \frac{\left(\frac{Y^{\perp\prime}Z^{\perp}}{\sqrt{T}}\frac{Z^{\perp\prime}u^{\perp}}{\sqrt{T}}\right)^2}{\frac{Y^{\perp\prime}Z^{\perp}}{\sqrt{T}}\widehat{S}\frac{Z^{\perp\prime}Y^{\perp}}{\sqrt{T}}}\,. \tag{40}$$

Then, from Assumptions L_{Π} and HL it follows that

$$\frac{Y^{\perp\prime}Z^{\perp}}{\sqrt{T}} = \frac{\Pi' Z^{\perp\prime}Z^{\perp}}{\sqrt{T}} + \frac{v_2' Z^{\perp}}{\sqrt{T}} = C'\frac{Z^{\perp\prime}Z^{\perp}}{T} + \frac{v_2' Z^{\perp}}{\sqrt{T}} \xrightarrow[d]{} (C+J)'\,, \tag{41}$$

$$\frac{Z^{\perp\prime}u^{\perp}}{\sqrt{T}} = \frac{Z^{\perp\prime}v_1}{\sqrt{T}} - \beta\frac{Z^{\perp\prime}v_2}{\sqrt{T}} \xrightarrow[d]{} K - \beta J\,. \tag{42}$$

If β was consistently estimable, then an appropriate HAC estimator (see e.g. Andrews, 1991) could consistently estimate S. However, this is not the case under weak instrument asymptotics. Let us adopt a general HAC estimator with kernel $k\left(\frac{j}{b(T)}\right)$, where $b(T)$ denotes the data-dependent bandwidth parameter. Then $\widehat{S} = \sum_{j=-T+1}^{T-1} k\left(\frac{j}{b(T)}\right)\widehat{\Gamma}_j$, where $\widehat{\Gamma}_j \equiv T^{-1}\sum_{t=j+1}^{T}\widehat{g}_t\widehat{g}_{t-j}'$ and $\widehat{g}_t \equiv Z_t^{\perp}\widehat{u}_t = Z_t^{\perp}u_t^{\perp} - \left(\widehat{\beta}-\beta\right)Z_t^{\perp}Y_t^{\perp}$. Let

$$\Xi_j \equiv E\left[\begin{pmatrix} Z_t^{\perp}v_{1,t} \\ Z_t^{\perp}v_{2,t} \end{pmatrix}\begin{pmatrix} Z_{t-j}^{\perp\prime}v_{1,t-j} & Z_{t-j}^{\perp\prime}v_{2,t-j} \end{pmatrix}\right] = \begin{bmatrix} \Xi_j^{11} & \Xi_j^{12} \\ \Xi_j^{21} & \Xi_j^{22} \end{bmatrix}, \tag{43}$$

where the sub-matrices of Ξ_j are all $(K_2 \times K_2)$ square matrices. Furthermore,

$$\widetilde{B}_{T,j} = T^{-1}\sum Z_t^{\perp}u_t^{\perp}Z_{t-j}^{\perp\prime}u_{t-j}^{\perp} \underset{p}{\rightarrow} \Xi_j^{11} - \beta\Xi_j^{12} - \beta\Xi_j^{21} + \beta^2\Xi_j^{22}, \tag{44}$$

$$\widetilde{F}_{T,j} = T^{-1}\sum Z_t^{\perp}u_t^{\perp}Z_{t-j}^{\perp\prime}Y_{t-j}^{\perp} + T^{-1}\sum Z_t^{\perp}Y_t^{\perp}Z_{t-j}^{\perp\prime}u_{t-j}^{\perp} \underset{p}{\rightarrow} \Xi_j^{12} - 2\beta\Xi_j^{22} + \Xi_j^{21}\ , \tag{45}$$

$$\widetilde{D}_{T,j} = T^{-1}\sum Z_t^{\perp}Y_t^{\perp}Z_{t-j}^{\perp\prime}Y_{t-j}^{\perp} \underset{p}{\rightarrow} \Xi_j^{22}, \tag{46}$$

$$\widehat{\beta}-\beta = \left(\frac{Y^{\perp\prime}Z^{\perp}}{\sqrt{T}}\frac{Z^{\perp\prime}Y^{\perp}}{\sqrt{T}}\right)^{-1}\left(\frac{Y^{\perp\prime}Z^{\perp}}{\sqrt{T}}\frac{Z^{\perp\prime}u^{\perp}}{\sqrt{T}}\right) \underset{d}{\rightarrow} \frac{(C+J)'(K-\beta J)}{(C+J)'(C+J)}, \tag{47}$$

where the summations go from $t = j+1$ to T. Hence we can write $\widehat{\Gamma}_j = \widetilde{B}_{T,j} - \left(\widehat{\beta}-\beta\right)\widetilde{F}_{T,j} + \left(\widehat{\beta}-\beta\right)^2\widetilde{D}_{T,j}$. From these, the asymptotic distribution of the Wald statistic follows, denoted by $\mathbb{W}_{\infty}\left(\beta, C, \{\Xi_j\}\right)$, where the notation highlights the dependence on the arguments: the consistently estimable entries of Ξ_j, and C and β. Importantly, $\mathbb{W}_{\infty}(\cdot)$ will not depend on the particular kernel function $k(\cdot)$ provided that standard regularity conditions (see e.g. Andrews, 1991) are satisfied.

The asymptotic size distortion of the Wald test with nominal size ν is defined as

$$s_{\mathbb{W}_{\infty},\nu}\left(\beta, C, \{\Xi_j\}\right) \equiv \mathrm{P}\left(\mathbb{W}_{\infty}\left(\beta, C, \{\Xi_j\}\right) > \chi^2_{1,1-\nu}\right) - \nu\,. \tag{48}$$

For β and C, we can construct a $(1-\alpha)$ level joint confidence interval using eq. (34). While we are not aware of a simple expression for the distribution of $\mathbb{W}_{\infty}$, eq. (48) can be easily simulated for any $(\widetilde{C},\widetilde{\beta}) \in \mathrm{CI}_{1-\alpha}^{C,\beta}$ and using the point estimates of the Ξ_js, denoted by $\widehat{\Xi}_j$.

Define $L_{1-\alpha}^{s} \equiv \min\limits_{(\widetilde{C},\widetilde{\beta})\in\mathrm{CI}_{1-\alpha}^{C,\beta}} s_{\mathbb{W}_{\infty},\nu}(\widetilde{\beta},\widetilde{C},\left\{\widehat{\Xi}_j\right\})$ and $U_{1-\alpha}^{s} \equiv \max\limits_{(\widetilde{C},\widetilde{\beta})\in\mathrm{CI}_{1-\alpha}^{C,\beta}} s_{\mathbb{W}_{\infty},\nu}(\widetilde{\beta},\widetilde{C},\left\{\widehat{\Xi}_j\right\})$. Our proposed $(1-\alpha)$ level asymptotic confidence interval for $s_{\mathbb{W}_{\infty},\nu}\left(\beta, C, \{\Xi_j\}\right)$ is

$$\mathrm{CI}_{1-\alpha}^{s_{\mathbb{W}_{\infty},\nu}} = \begin{cases} \left[L_{1-\alpha}^{s}, U_{1-\alpha}^{s}\right] & \text{if } \mathrm{CI}_{1-\alpha}^{C,\beta} \neq \emptyset\,, \\ \emptyset & \text{if } \mathrm{CI}_{1-\alpha}^{C,\beta} = \emptyset\,. \end{cases} \tag{49}$$

Proposition 3 (Confidence interval for the Wald size distortion): Under Assumptions

L_{Π} and HL, $\text{CI}_{1-\alpha}^{s_{\mathbb{W}_{\infty},\nu}}$ in eq. (49) is an asymptotically valid confidence interval for the Wald size distortion $s_{\mathbb{W}_{\infty},\nu}\left(\beta, C, \{\Xi_j\}\right)$, that is

$$\lim_{T\to\infty} \text{P}\left(s_{\mathbb{W}_{\infty},\nu}\left(\beta, C, \{\Xi_j\}\right) \in \text{CI}_{1-\alpha}^{s_{\mathbb{W}_{\infty},\nu}}\right) \geq 1-\alpha\,. \tag{50}$$

Proof. See Section A of the Online Appendix. □

3.3 The Local Projections–IV Method

Since Jordà (2005), the local projections method has become popular 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) is used in several studies, e.g. Jordà et al. (2015) and Ramey and Zubairy (2018). Stock and Watson (2018) provides an overview of the LP–IV econometric framework, which we adopt.

Consider the $(k \times 1)$ 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, $\Theta(L) = \Theta_0 + \Theta_1 L + \Theta_2 L^2 + \ldots$, and Θ_h is a $(k \times m)$ matrix of coefficients. Furthermore, ε_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 Θ_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 $Y_{1,t} = \varepsilon_{1,t} + \{\varepsilon_{2:m,t}, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$, where $\varepsilon_{2:m,t} \equiv (\varepsilon_{2,t}, \varepsilon_{3,t}, \ldots \varepsilon_{m,t})'$, and the shorthand $\{\cdot\}$ denotes the linear combination of the variables inside the braces. 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 $\Theta_{h,i1}$ in the regression:

$$Y_{i,t+h} = \Theta_{h,i1} Y_{1,t} + u_{i,t+h}^{h}\,, \tag{51}$$

where $u_{i,t+h}^{h} = \{\varepsilon_{t+h}, \ldots, \varepsilon_{t+1}, \varepsilon_{2:m,t}, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$. 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. (51), 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

$$Y_{i,t+h}^{\perp} = \Theta_{h,i1} Y_{1,t}^{\perp} + u_{i,t+h}^{h\perp}\,. \tag{52}$$

As Stock and Watson (2018) note, the control variables can serve two purposes: first, the exogeneity conditions $\mathrm{E}(\varepsilon_{2:m,t}Z_t') = 0$, and $\mathrm{E}(\varepsilon_{t+j}Z_t') = 0$ 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 $\mathrm{E}(\varepsilon_{2:m,t}^{\perp}Z_t^{\perp\prime}) = 0$, and $\mathrm{E}(\varepsilon_{t+j}^{\perp}Z_t^{\perp\prime}) = 0$ for all $j \neq 0$. Instrument relevance is given by $\mathrm{E}(\varepsilon_{1,t}^{\perp}Z_t^{\perp}) = \Pi$. Note that under instrument exogeneity, the instrument relevance condition is equivalent to $\mathrm{E}(Y_{1,t}^{\perp}Z_t^{\perp}) = \Pi$, which suggests the familiar first stage equation (using the same normalization $Z^{\perp\prime}Z^{\perp}/T = I_{K_2}$ as before, and $Y_{1,t}^{\perp}$ acting as the endogenous regressor $Y_t^{\perp}$):

$$Y_{1,t}^{\perp} = Z_t^{\perp\prime}\Pi + v_{2,t}\,, \tag{53}$$

where $\mathrm{E}(v_{2,t}Z_t^{\perp}) = 0$. The reduced-form structural equation is:

$$Y_{i,t+h}^{\perp} = Z_t^{\perp\prime}\Pi\Theta_{h,i1} + \Theta_{h,i1}v_{2,t} + u_{i,t+h}^{h\perp} = Z_t^{\perp\prime}\Pi\Theta_{h,i1} + v_{1,t}\,, \tag{54}$$

where $Y_{i,t+h}^{\perp}$ corresponds to y_t, $\Theta_{h,i1}$ plays the role of β, 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 $Z_t^{\perp\prime}v_{1,t}$ and $Z_t^{\perp\prime}v_{2,t}$ feature conditional heteroskedasticity and autocorrelation, hence our confidence interval in the previous subsections apply directly to the LP–IV framework under Assumptions L_{Π}, HL, instrument exogeneity, and the validity of the structural vector moving average representation described at the beginning of this subsection.

3.4 The General Case of Multiple Endogenous Regressors ($n \geq 1$)

The concentration matrix is the generalization of the concentration parameter $\mu_{K_2}^2$ in Section 3.2.1, and plays an analogous role in determining the strength of identification in terms of maximal bias and size distortion in the homoskedastic IV model.

Define $\lambda \equiv \Omega^{1/2}C\Sigma_{VV}^{-1/2}$. Then the concentration matrix Λ is given by

$$\Lambda \equiv \frac{1}{K_2}\Sigma_{VV}^{-1/2\prime}C'\Omega C\Sigma_{VV}^{-1/2} = \frac{1}{K_2}\lambda'\lambda\,. \tag{55}$$

In the general case of $n \geq 1$ endogenous regressors, the worst-case bias and size distortion are functions of the minimum eigenvalue of the concentration matrix, denoted by $\mathrm{mineval}\,(\Lambda)$, as shown by Stock and Yogo (2005).

Similarly to the previously discussed case of $n = 1$, our proposed confidence interval builds on the asymptotic distribution of the OLS estimator of Π in eq. (13):

$$\sqrt{T}\left(\widehat{\Pi}_T - \Pi\right) = \left(T^{-1}Z^{\perp\prime}Z^{\perp}\right)^{-1}T^{-1/2}Z^{\perp\prime}V^{\perp}\,, \tag{56}$$

$$\sqrt{T}\,\text{vec}\left(\widehat{\Pi}_T - \Pi\right) \underset{d}{\rightarrow} \mathcal{N}\left(0, \Sigma_{VV} \otimes \Omega^{-1}\right), \tag{57}$$

$$\text{vec}\left(\widehat{C} - C\right) \underset{d}{\rightarrow} \mathcal{N}\left(0, \Sigma_{VV} \otimes \Omega^{-1}\right), \tag{58}$$

where eq. (57) follows directly from eq. (56) and Assumption M, and in eq. (58) we used $\Pi = \Pi_T = C/\sqrt{T}$ and $\widehat{C} \equiv \widehat{\Pi}_T\sqrt{T}$. While $\widehat{C}$ is an inconsistent estimator of C, for our purposes the asymptotic normality result of eq. (58) is sufficient. Note that $\left[\text{vec}\left(\widehat{C} - C\right)\right]' \left[\Sigma_{VV} \otimes \Omega^{-1}\right]^{-1} \left[\text{vec}\left(\widehat{C} - C\right)\right] \underset{d}{\rightarrow} \chi^2_{nK_2}$. By using $\widehat{\Sigma}_{VV} \underset{p}{\rightarrow} \Sigma_{VV}$ and $\widehat{\Omega} \equiv Z^{\perp\prime}Z^{\perp}/T \underset{p}{\rightarrow} \Omega$, we obtain the distribution of the Wald statistic, $\mathcal{W}(C)$:

$$\mathcal{W}(C) \equiv \left[\text{vec}\left(\widehat{C} - C\right)\right]' \left[\widehat{\Sigma}_{VV} \otimes \widehat{\Omega}^{-1}\right]^{-1} \left[\text{vec}\left(\widehat{C} - C\right)\right] \underset{d}{\rightarrow} \chi^2_{nK_2}. \tag{59}$$

By taking the $(1-\alpha)$ quantile of the $\chi^2_{nK_2}$ distribution (denoted by $\chi^2_{nK_2,1-\alpha}$), $\mathcal{W}(C)$ can be inverted to obtain an asymptotically valid $(1-\alpha)$ level confidence set for C:

$$\text{CI}^C_{1-\alpha} \equiv \left\{\forall\, \widetilde{C} \in \mathbb{R}^{K_2 \times n} : \mathcal{W}\left(\widetilde{C}\right) \le \chi^2_{nK_2,1-\alpha}\right\}. \tag{60}$$

Note that $\text{CI}^C_{1-\alpha}$ is compact and non-empty. Using the definition of Λ (eq. (55)), define

$$\widetilde{\Lambda}(\widetilde{C}) \equiv \frac{1}{K_2}\widehat{\Sigma}_{VV}^{-1/2\prime}\widetilde{C}'\widehat{\Omega}\widetilde{C}\widehat{\Sigma}_{VV}^{-1/2}, \tag{61}$$

which is a continuous function of $\widetilde{C}$ and of the consistent estimates of Σ_{VV} and Ω. Let

$$L^{\Lambda}_{1-\alpha} \equiv \min_{\widetilde{C} \in \text{CI}^C_{1-\alpha}} \text{mineval}(\widetilde{\Lambda}(\widetilde{C})) \qquad U^{\Lambda}_{1-\alpha} \equiv \max_{\widetilde{C} \in \text{CI}^C_{1-\alpha}} \text{mineval}(\widetilde{\Lambda}(\widetilde{C})). \tag{62}$$

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

$$\text{CI}^{\Lambda}_{1-\alpha} \equiv \left[L^{\Lambda}_{1-\alpha}, U^{\Lambda}_{1-\alpha}\right]. \tag{63}$$

Furthermore, let us define

$$L^b_{1-\alpha} \equiv b(U^{\Lambda}_{1-\alpha}; n, K_2) \qquad U^b_{1-\alpha} \equiv b(L^{\Lambda}_{1-\alpha}; n, K_2), \tag{64}$$

$$L^s_{1-\alpha} \equiv s(U^{\Lambda}_{1-\alpha}; n, K_2) \qquad U^s_{1-\alpha} \equiv s(L^{\Lambda}_{1-\alpha}; n, K_2), \tag{65}$$

which constitute the endpoints of the $(1-\alpha)$ level asymptotic confidence intervals for bias (eq. (64)) and size distortion (eq. (65)), as summarized in Proposition 4.

Proposition 4 (Confidence interval based on the projection method): Under Assumptions L_{Π} and M, $\text{CI}^{\Lambda}_{1-\alpha}$ is an asymptotically valid $(1-\alpha)$ level confidence interval for mineval (Λ), that is,

$$\lim_{T\to\infty} P\left(\text{mineval}\,(\Lambda) \in \text{CI}_{1-\alpha}^{\Lambda}\right) \geq 1-\alpha\,. \tag{66}$$

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:

$$\lim_{T\to\infty} P\left(b\,(\text{mineval}\,(\Lambda)\,; n, K_2) \in [L_{1-\alpha}^{b}, U_{1-\alpha}^{b}]\right) \geq 1-\alpha\,, \tag{67}$$

$$\lim_{T\to\infty} P\Big(s\,(\text{mineval}\,(\Lambda)\,; n, K_2) \in [L_{1-\alpha}^{s}, U_{1-\alpha}^{s}]\Big) \geq 1-\alpha\,. \tag{68}$$

Proof. See Section A of the Online Appendix. □

Remark 2. Note that, as $\widetilde{\Lambda}\left(\widetilde{C}\right)$ is not a one-to-one function of $\widetilde{C}$ in general, our proposed confidence interval is conservative. Hence, in the case of $n=1$ endogenous variable, we recommend using the confidence interval introduced in section 3.2.1, based on the non-central χ^2 distribution.

Remark 3. When there is only one endogenous regressor ($n=1$), then the Karush–Kuhn–Tucker conditions provide an analytical solution to eq. (62), and hence to eqs. (64) and (65) (we thank an anonymous referee for pointing this out). Define $d_T \equiv \widehat{\sigma}_{VV}^{-1}\widehat{C}'\widehat{\Omega}\widehat{C}$. The upper bound is given by $U_{1-\alpha}^{\Lambda} = K_2^{-1}\left(\sqrt{\chi^2_{K_2,1-\alpha}} + \sqrt{d_T}\right)^2$. If $d_T \geq \chi^2_{K_2,1-\alpha}$, then $L_{1-\alpha}^{\Lambda} = K_2^{-1}\left(\sqrt{d_T} - \sqrt{\chi^2_{K_2,1-\alpha}}\right)^2$, while if $d_T < \chi^2_{K_2,1-\alpha}$, 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.

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 $Y = Z\Pi + V$, 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. In the homoskedastic DGP (Data Generating Process) we specify $V_t \sim$ *iid* $\mathcal{N}(0, I_n)$ and $Z_t \sim$ *iid* $\mathcal{N}(0, I_{K_2})$, and consider $n = 1$, with $K_2 = \{n+1, \ldots, n+4\}$ when focusing on bias, and $K_2 = \{n, \ldots, n+3\}$ 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 $T = \{100, 250, 500, 1000\}$. We constructed the confidence intervals based on both the non-central chi-squared approach of Proposition 1 and the projection method of Proposition 4. 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, mineval(Λ), K_2, and n. In the interest of space, we relegated the Monte Carlo results for $n = 2$ to Section C of the Online Appendix.

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 $\widetilde{\Lambda}(\widetilde{C})$ 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 $\widetilde{\Lambda}(\widetilde{C})$ 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-squared approximation.

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

4.2 The Heteroskedastic/Autocorrelated IV Model

We consider two DGPs, labeled as DGP 1N and DGP 2N, to construct confidence intervals for the Nagar bias defined in eq. (27), and DGP 1S and DGP 2S to construct confidence intervals for the size distortion of the Wald test at the $\nu = 5\%$ nominal level. DGPs 1N and 2N are inspired by Montiel Olea and Pflueger (2013, p. 361), and

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

Panel A. Coverage rates (non-central χ^2)

	$K_2 = 2$			$K_2 = 3$			$K_2 = 4$			$K_2 = 5$		
$T \backslash b$	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

Panel B. Coverage rates (projection method)

	$K_2 = 2$			$K_2 = 3$			$K_2 = 4$			$K_2 = 5$		
$T \backslash b$	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

Panel C. Median lengths of confidence intervals (non-central χ^2)

	$K_2 = 2$			$K_2 = 3$			$K_2 = 4$			$K_2 = 5$		
$T \backslash b$	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

Panel D. Median lengths of confidence intervals (projection method)

	$K_2 = 2$			$K_2 = 3$			$K_2 = 4$			$K_2 = 5$		
$T \backslash b$	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 χ^2 approximation for different sample sizes T, values of b, and numbers of instruments K_2 in 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 $(1 - \alpha) = 0.90$.

feature conditional heteroskedasticity but no autocorrelation, while DGPs 2N and 2S have both. First we describe DGPs 2N and 2S, and then discuss the restriction under which we obtain DGPs 1N and 1S.

Let $\widetilde{Z}_t = \left(\widetilde{Z}_{1,t}, \ldots, \widetilde{Z}_{K_2,t}\right)'$, $\epsilon_t \sim$ *iid* $\mathcal{N}(0, (1-\rho^2) I_{K_2})$ and $\widetilde{Z}_t = \rho \widetilde{Z}_{t-1} + \epsilon_t$, where ρ controls the persistence of the independent autoregressive processes. The $(T \times K_2)$ matrix $\widetilde{Z}$ collects the vectors $\widetilde{Z}_t$. Let Z_t be the standardized $\widetilde{Z}_t$ in-sample, such that $Z'Z/T = I_{K_2}$. That is, let Z^{std} be the (column-by-column) demeaned $\widetilde{Z}$ divided by its standard deviation (column-by-column), and $Q_Z^{\text{std}} = (Z^{\text{std}\prime} Z^{\text{std}}/T)^{-1/2}$. Then

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

Panel A. Coverage rates (non-central χ^2)

	$K_2 = 1$			$K_2 = 2$			$K_2 = 3$			$K_2 = 4$		
$T \backslash s$	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

Panel B. Coverage rates (projection method)

	$K_2 = 1$			$K_2 = 2$			$K_2 = 3$			$K_2 = 4$		
$T \backslash s$	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

Panel C. Median lengths of confidence intervals (non-central χ^2)

	$K_2 = 1$			$K_2 = 2$			$K_2 = 3$			$K_2 = 4$		
$T \backslash s$	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

Panel D. Median lengths of confidence intervals (projection method)

	$K_2 = 1$			$K_2 = 2$			$K_2 = 3$			$K_2 = 4$		
$T \backslash s$	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 χ^2 approximation for different sample sizes T, values of s, and numbers of instruments K_2 in 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 $(1 - \alpha) = 0.90$.

$Z = Z^{\text{std}} Q_Z^{\text{std}}$. We specify a moving average process $u_{2t} = q_t + \theta_1 q_{t-1}$, where $q_t \sim$ *iid* $\mathcal{N}(0, \sigma_q^2)$, and it is independent of $\widetilde{Z}_t$ (and Z_t) both contemporaneously and at all leads and lags. Furthermore, let $b_t \sim$ *iid* $\mathcal{N}(0, \sigma_b^2)$ (independent of all the previous random variables both contemporaneously and at all leads and lags), and $u_{1t} = \theta_2 u_{2t} + b_t$, where $\theta_2 = \tilde{\alpha}/(1 + \theta_1^2)$. Define the $(K_2 \times 1)$ coefficient vectors $\widetilde{\gamma}_1$ as $\widetilde{\gamma}_1 = (\gamma_1, 0, \dots, 0)'$ and $\widetilde{\gamma}_2 = (\gamma_2, 0, \dots, 0)'$. Conditional heteroskedasticity is introduced by letting $v_{1t} = Z_t' \widetilde{\gamma}_1 u_{1t}$ and $v_{2t} = Z_t' \widetilde{\gamma}_2 u_{2t}$ be the t-th element of v_1 and v_2, respectively.

In DGPs 1N and 1S, we set $\theta_1 = 0$ to make $(Z_t' v_{1t}, Z_t' v_{2t})'$ serially uncorrelated, while preserving conditional heteroskedasticity. Section C of the Online Appendix presents the details of the simulation design.

For the Nagar bias, 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, $N_{\text{TSLS}}(\beta, C, W) = \{0.05, 0.10, 0.3\}$ (in the case of $K_2 = 4$ instruments, we specified $N_{\text{TSLS}}(\beta, C, W) = \{-0.05, -0.10, -0.3\}$, as the Nagar bias is non-positive in this case). For the Wald size distortion, we performed the simulations for the just-identified case. We set $\beta = 1$ in all cases. Furthermore, we specified $C = \left(c_*^2, c_*, \ldots, c_*\right)$, and using Matlab's `fzero` or `fsolve` solver we determined the value of c_* such that (given the rest of the parameters) it implies the desired amount of Nagar bias or size distortion. The specific C vectors, along with the derivation of the covariance matrix W and details of the numerical optimization can be found in Section C of the Online Appendix. In the cases of DGPs 1N and 1S, we used the White (1980) covariance estimator, while for DGPs 2N and 2S, we used a rectangular kernel with one lag (in the rare cases when it delivered non-positive definite estimates, we replaced them by the Newey and West (1987) estimates).

Results for the Nagar bias are reported in Table 4, showing that our proposed confidence interval delivers valid (although conservative) coverage rates in both DGPs across different values of Nagar bias $N_{\text{TSLS}}(\beta, C, W)$, sample sizes T, and numbers of instruments K_2 .

Table 4: Coverage rates for the Nagar bias

Panel A. Heteroskedastic IV model												
	$K_2 = 1$			$K_2 = 2$			$K_2 = 3$			$K_2 = 4$		
$T \backslash N_{\text{TSLS}}$	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
Panel B. Heteroskedastic and autocorrelated IV model												
	$K_2 = 1$			$K_2 = 2$			$K_2 = 3$			$K_2 = 4$		
$T \backslash N_{\text{TSLS}}$	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 numbers of instruments K_2 in DGP 1N in Section 4.2. The lower panel displays analogous results, based on DGP 2N. The number of Monte Carlo simulations is 2000. The nominal coverage level is $(1 - \alpha) = 0.90$.

Table 5: Coverage rates for the Wald size distortion

$T \backslash s_{\mathbb{W}_\infty,\nu}$	Heteroskedastic IV model			Heteroskedastic and autocorrelated IV model		
	0.3	0.1	0.05	0.3	0.1	0.05
100	0.96	0.96	0.94	0.95	0.91	0.85
250	0.97	0.98	0.96	0.96	0.95	0.92
500	0.96	0.98	0.96	0.97	0.96	0.94
1000	0.97	0.98	0.97	0.97	0.96	0.95

Note: The table shows the empirical coverage rates of the proposed confidence interval for the Wald size distortion $s_{\mathbb{W}_\infty,\nu}\left(\beta, C, \{\Xi_j\}\right)$ for different sample sizes T and values of the size distortion (at the $\nu = 5\%$ nominal level). DGP 1S is heteroskedastic, while DGP 2S features both heteroskedasticity and autocorrelation. 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 consumption Euler equation, as in Yogo (2004) and Montiel Olea and Pflueger (2013). The model is a linear IV model; we consider both homoskedastic, and heteroskedastic/serially correlated cases.

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

$$\Delta c_{t+1} = \nu + \psi r_{t+1} + u_{t+1}\,, \tag{69}$$

$$r_{t+1} = \xi + \psi^{-1}\Delta c_{t+1} + \eta_{t+1}\,, \tag{70}$$

where Δ denotes the first difference operator, c_t is the logarithm of the level of consumption, Δc_{t+1} is consumption growth, and r_{t+1} is a real asset return, ψ is the IES parameter, ν and ξ are constants, while u_{t+1} and η_{t+1} are stochastic disturbances, which can be conditionally heteroskedastic or autocorrelated. Note that eq. (70) expresses the same relationship between consumption growth and returns as eq. (69), but often the estimates of ψ 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. (69) and (70), we use real per capita consumption growth for Δc_{t+1}, and the real return on the short-term interest rate for r_{t+1}. As Yogo (2004) notes, by using instruments dated $t-1$, ψ or its reciprocal ψ^{-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 in the left and right panels, the point estimates suggest contradicting values for ψ, an empirical result also emphasized by Yogo (2004) and Montiel Olea and Pflueger (2013).

Furthermore, for the specification in eq. (70) 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. (69): 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. Our results confirm the latter authors' finding, that the test developed for the homoskedastic case can be misleading in the presence of heteroskedasticity or autocorrelation. Our confidence intervals for the Wald size distortion $\mathrm{CI}_{0.95}^{\mathbb{S}\mathbb{W}_{\infty,\nu}}$ signal strong identification in eq. (69) but weak identification in eq. (70). 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 meta analysis of Havránek (2015), who finds that, after correcting for publication bias, IES estimates based on macroeconomic data are centered around zero.

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.

Table 6: Intertemporal elasticity of substitution

Panel A. Heteroskedastic/serially correlated IV model	IES ψ	IES ψ^{-1}
TSLS estimate (standard error)	0.06 (0.098)	0.68 (0.813)
$\mathrm{CI}_{0.95}^{N_{\mathrm{TSLS}}}$	$[-0.00, 0.02]$	$[20.28, 12716.19]$
$\mathrm{CI}_{0.95}^{S_{W_\infty,\nu}}$	$[-0.01, 0.01]$	$[0.10, 0.95]$
$\widehat{F}_{\mathrm{eff}}$	8.14	2.65
Critical value ($\tau = 0.1$)	15.49	13.99
Critical value ($\tau = 0.3$)	**7.75**	7.04
Panel B. Homoskedastic IV model	IES ψ	IES ψ^{-1}
TSLS 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	$[0.033, 0.089]$	$[0.105, 0.822]$
F-statistic	15.53	2.93
Critical value (5% bias)	16.85	16.85
Critical value (10% bias)	**10.27**	10.27
Critical value (5% size distortion)	24.58	24.58
Critical value (10% size distortion)	**13.96**	13.96

Note: The table displays the estimation results of the consumption Euler equations with Δc_{t+1} regressed on r_{t+1} (specification IES ψ in eq. (69)), and r_{t+1} regressed on Δc_{t+1} (specification IES ψ^{-1} in eq. (70)). Panel A shows results based on the heteroskedastic and autocorrelated IV model: TSLS 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 intervals for the Nagar bias and the Wald size distortion (nominal level: 5%, number of simulations: 10,000), along with the effective F-statistics $\widehat{F}_{\mathrm{eff}}$ and the corresponding 5% critical values, allowing for τ 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 χ^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.

The structural equations for the ZLB, normal and linear specifications are:

$$\begin{aligned}\sum_{j=0}^{h} y_{t+j} = c_h^{\mathrm{ZLB}} + \gamma_h^{\mathrm{ZLB}} I_{t-1} + I_{t-1}\phi_h^{\mathrm{ZLB}}(L) z_{t-1} + (1 - I_{t-1})\, \xi_h^{\mathrm{ZLB}}(L) z_{t-1} \\ + m_h^{\mathrm{ZLB}} \sum_{j=0}^{h} g_{t+j} I_{t-1} + \omega_{t+h}^{\mathrm{ZLB}},\end{aligned} \tag{71}$$

$$\begin{aligned}\sum_{j=0}^{h} y_{t+j} = c_h^{\mathrm{normal}} + \gamma_h^{\mathrm{normal}} I_{t-1} + I_{t-1}\phi_h^{\mathrm{normal}}(L) z_{t-1} + (1 - I_{t-1})\, \xi_h^{\mathrm{normal}}(L) z_{t-1} \\ + m_h^{\mathrm{normal}} \sum_{j=0}^{h} g_{t+j} (1 - I_{t-1}) + \omega_{t+h}^{\mathrm{normal}},\end{aligned} \tag{72}$$

$$\sum_{j=0}^{h} y_{t+j} = c_h^{\mathrm{linear}} + \phi_h^{\mathrm{linear}}(L) z_{t-1} + m_h^{\mathrm{linear}} \sum_{j=0}^{h} g_{t+j} + \omega_{t+h}^{\mathrm{linear}}, \tag{73}$$

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, γ_h^s are scalar coefficients; $\phi_h^s(L)$ and $\zeta_h^s(L)$ are polynomials in the lag operator L ($L=0,1,2,3$); m_h^s are the government spending multipliers, which are the structural parameters of interest. The error terms ω_{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. In the ZLB and normal specifications, the instruments are multiplied by the appropriate indicator.

Table 7: Government spending multipliers

IV(s)	Horizon	Estimates				CIs for bias and size distortion			$\widehat{F}_{\text{eff}}$ 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)	N:	$[0.00, 0.13]$	$[0.00, 0.12]$	$[0.00, 0.04]$	19.95 [23.11]	22.61 [23.11]	**43.68** [23.11]
					S:	$[-0.01, 0.06]$	$[-0.02, 0.04]$	$[-0.02, -0.00]$			
	4Y	0.71 (0.04)	0.77 (0.06)	0.77 (0.38)	N:	$[0.00, 0.22]$	$[0.00, 0.54]$	$[-0.04, 0.20]$	11.55 [23.11]	10.21 [23.11]	**24.06** [23.11]
					S:	$[-0.05, 0.09]$	$[-0.05, 0.13]$	$[-0.04, 0.01]$			
BP	2Y	0.38 (0.11)	0.64 (0.03)	0.10 (0.11)	N:	$[-0.07, -0.01]$	$[-0.00, 0.00]$	$[-0.01, 0.00]$	**36.72** [23.11]	**53.98** [23.11]	**70.60** [23.11]
					S:	$[-0.00, 0.02]$	$[-0.02, -0.00]$	$[-0.01, -0.00]$			
	4Y	0.47 (0.11)	0.71 (0.03)	0.12 (0.12)	N:	$[-0.21, -0.00]$	$[-0.00, 0.02]$	$[-0.02, 0.01]$	20.11 [23.11]	21.03 [23.11]	**36.44** [23.11]
					S:	$[-0.01, 0.05]$	$[-0.04, 0.01]$	$[-0.03, -0.00]$			
News & BP	2Y	0.42 (0.10)	0.67 (0.03)	0.26 (0.10)	N:	$\emptyset$	$[-0.00, 0.00]$	$[-0.00, -0.00]$	**37.85** [13.19]	**37.20** [13.56]	**37.99** [13.06]
					S:	$\emptyset$	$[-0.01, 0.00]$	$[-0.01, -0.01]$			
	4Y	0.56 (0.08)	0.76 (0.04)	0.21 (0.14)	N:	$\emptyset$	$[-0.01, 0.00]$	$[-0.07, 0.00]$	14.90 [15.46]	12.11 [15.89]	**19.43** [18.20]
					S:	$\emptyset$	$[-0.03, 0.03]$	$[-0.02, 0.03]$			

Note: The columns labeled "Estimates" report TSLS 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 "CIs for bias and size distortion" report the 95% confidence intervals for the Nagar bias ("N:") and the Wald size distortion ("S:", nominal level: 5%, number of simulations: 10,000). 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.

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 "CIs for bias and size distortion" report the confidence intervals for the Nagar bias of the TSLS estimator (rows labeled "N:") and the corresponding Wald test's size distortion (rows labeled "S:"). 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 $\tau = 0.1$ maximum relative bias in brackets.

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 and size distortions when estimating the state-dependent model, either using the BP or both the BP and the news shocks as instruments, while we find some positive bias and non-negligible size distortion when using only the news shock instrument. After correcting the TSLS 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. In general, the confidence intervals for size distortion lead to similar conclusions. These results demonstrate that our proposed confidence can indeed provide additional and useful information.

Turning to the linear, state-independent specification of the model in eq. (73), 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 BP instrument. However, when using both the BP 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 $\mathrm{CI}_{1-\alpha}^{\widetilde{C},\beta}$ is empty, meaning there is no $(\widetilde{C}, \widetilde{\beta}) \in \mathbb{R}^3$ 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 Appendix contains the proofs, further theoretical and Monte Carlo results, and the description of the data sets used in the present paper. Replication code is available on the journal's website.

ACKNOWLEDGMENTS

We are grateful to Todd Clark and two anonymous referees for very useful comments. 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. 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 Economy and Competitiveness through the Severo Ochoa Programme for Centers of Excellence in R&D (SEV-2015-0563). The views expressed herein are those of the authors and should not be attributed to the Banco de España or the Eurosystem.

REFERENCES

Andrews, D. W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817-858.

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.

Andrews, I., Stock, J. H., and Sun, L. (forthcoming), "Weak Instruments in IV Regression: Theory and Practice," *Annual Review of Economics*.

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

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.

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.

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.

Montiel Olea, J. L., and Pflueger, C. (2013), "A Robust Test for Weak Instruments," *Journal of Business & Economic Statistics*, 31, 358-369.

Montiel Olea, J. L., Stock, J. H., and Watson M. W. (2018), "Inference in Structural Vector Autoregressions with External Instruments," *Working Paper*.

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.

White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.

Yogo, M. (2004), "Estimating the Elasticity of Intertemporal Substitution When Instruments Are Weak," *The Review of Economics and Statistics*, 86, 797-810.