Alma Mater Studiorum
Università di Bologna

DOTTORATO DI RICERCA IN
SCIENZE STATISTICHE

CICLO XXXI

Settore Concorsuale:	13/A5
Settore Scientifico Disciplinare:	SECS-P/05

Essays on bootstrap inference under weakly identified models

Presentata da: Riccardo Ievoli

Coordinatrice Dottorato:
Prof.ssa Alessandra Luati

Supervisor:
Prof. Giuseppe Cavaliere

Esame Finale Anno 2019

Abstract

Instrumental Variables (IV) are widely used in econometrics to overcome endogeneity problem in regression models, which occurs when regressors are correlated with the stochastic component. Nonetheless, in applied works, practitioners face with instruments that are collectively "weak", i.e. poorly correlated with endogenous regressors. Under weak instruments, conventional estimators are no longer consistent and asymptotically normal. Furthermore, bootstrap methods could be useful to improve inference in IV estimation. However, under poorly relevant instruments, the bootstrap is deemed invalid and its use is generally discouraged in applied papers. In this work, we propose a new derivation of bootstrapped IV estimators under weak instruments asymptotics (Stock and Yogo, 2005) using residual–based bootstrap method involving fixed or resampled instruments. We prove that bootstrap counterpart of estimators, conditionally on the data, converges to a random distribution preserving some patterns (non–normality) of weak and irrelevant instruments scenarios. These issues may be also reflected in bootstrap–based confidence sets and hypothesis testing. In this sense, we explore the usefulness of bootstrap methods to provide information on the weakness (or the strength) of the instruments. We consider descriptive indicators and develop new bootstrap-based tests useful to detect weak instruments in IV framework. The method basically relies on Angelini et al. (2016) and allows to test normality of a certain number of (possibly standardized) bootstrap replications. Since conventional normality tests can lose power in presence of more instruments and high endogeneity, we propose new test statistics with the aim to test standard normality on the bootstrap replications. These tests are based on the moments of standard normal and are asymptotically chi–square distributed under the null hypothesis. In conclusion, we find that, in some cases, bootstrapped estimators may be used to test weak identification.

Contents

Chapter 1

Introduction

1.1 Instrumental variables and weak instruments

Since the early works of Nelson and Startz (1990a,b), it is well-known that finite sample distributions of conventional instrumental variable (IV) and two stage least squares (TSLS) estimators are asymptotically non–normal and severely biased under weak instruments, i.e. when instruments are not collectively correlated with the endogenous regressors. This usually lead to identification issues in the estimation of structural parameters. Therefore, t/Wald tests associated to IV/TSLS are generally oversized, providing unreliable and/or spurious inference, even in large samples, as pointed out by Bound et al. (1995). In the linear IV case, practitioners mainly adopt the first stage F rule of thumb ($F > 10$) or the F statistic with critical values developed by Stock and Yogo (2005), in order to determine if instruments are collectively weak or not. Nowadays this procedure is considered a standard tool but has its limitations, regarding mostly the fact that it is based on strictly assumptions and it does not consider the level of endogeneity, that crucially affects bias of TSLS as pointed out by Hahn and Hausman (2002a). To overcome these problems, different "robust", with respect to weak instruments, test statistics have been proposed (Kleibergen, 2002; Moreira, 2003), and several "partially robust" estimators, such as Limited Information Maximum Likelihood (LIML), may outperform conventional IV estimators, producing more reliable confidence intervals, as pointed out by Blomquist and Dahlberg (1999) among others.

Despite the available robust tests, evaluation of the instruments strength is still under discussion, especially under non–standard conditions and in presence of many endogenous regressors. In most non–standard cases the F statistic may be uninformative. Furthermore, there are arguments against robust methods, especially when partially robust estimation is conducted under very weak instruments, as explained by Hahn et al. (2004) and, more recenlty, Young (2017).

1.2 Bootstrap methods in weakly identified models

Bootstrap methods in econometrics are developed to improve the performance of estimators and test statistics, obtaining an approximation of sampling distributions, more precise standard errors and confidence intervals presenting better coverage rates. However, the bootstrap is not deemed valid, even in the linear IV framework, under weak instruments, especially when inference is conducted through non-robust IV estimators and associated t/Wald tests. This also appears in non–linear models, under the so called *weak identification* issue. The sources of bootstrap failure may be different, including the presence of outliers and incorrect resampling scheme, as examined by Canty et al. (2006). In IV models, this failure mostly occurs because the statistics of interest (as the t/Wald statistic) can be not asymptotically pivotal. In addition, as Horowitz (2001) pointed out, the bootstrap does not always perform well when the covariance matrix of certain coefficients is nearly singular as often appears under very weak (nearly irrelevant) instruments. Moreover, Moreira et al. (2009) prove the validity of bootstrap, even in the first order, for score and conditional likelihood ratio statistics under irrelevant and weak instruments. Furthermore, Dovonon and Gonçalves (2017) demonstrated the validity of bootstrap under local identification failures, testing overidentified restrictions in non–linear Generalized Method of the Moments (GMM). Nevertheless, there are some cases where both standard and new bootstrap methods remain invalid under weak instruments, as recently demonstrated by Doko Tchatoka (2015) and Wang and Kaffo (2016).

In the context of weakly identified models, the bootstrap is rarely applied to highlight non-normality and size distortion of the estimators. Flores-Lagunes (2007) exploits bootstrap as a bias–correction device for different IV estimators, showing its poor performance under weak instruments, while Ouysse (2011) finds similar results using a (fast) double bootstrap procedure. Zhan (2017) introduces bootstrap–based test to evaluate the maximal size distortion of the t–test under weak instruments, suggesting the informativness of graphical evaluation to compare the standardized bootstrap distribution (of an estimator) against standard normal. In the context of linearized rational expectation models, Angelini et al. (2016) develop a bootstrap–based approach to detect misspecification in dynamic stochastic general equilibrium (DSGE) models, testing normality directly on a certain (moderately small) number of bootstrap realizations. Caner (2011) suggests a bootstrap-based Kolmorov–Smirnov test between the Wald statistic and a chi–square distribution to distinguish between weak and nearly-weak identification in non-linear GMM inference. Furthermore, Camponovo and Otsu (2012) compare performances of bootstrap–based t/Wald test (in IV and TSLS estimation) in presence of outliers, arguing that performances can be dramatically worse in this case. Wang et al. (2015) propose the bootstrap in a selection method of the instrumental variables based on bootstrap version of approximate mean square error. Recently, Young (2017) applies the iterated bootstrap to study the distribution of estimators and tests of more than 1000 IV regressions from published papers, finding misleading results in terms of weak instruments tests, under iid theory, and checking understated confidence intervals. From a different point of view, Kitagawa (2015) proposes a bootstrap procedure to test exogeneity of instrument in the heterogeneous treatment effect model with binary endogenous regressor and discrete instrument.

From our perspective, invalidity of bootstrap methods could be used to detect weak instruments in IV models, in particular when two stage least squares estimator (TSLS) is applied. Firstly, the asymptotic distribution of TSLS estimators and associated t–test statistics is found to be substantially different from the normal distribution under weak instruments. Monte Carlo simulation, based on weak instrument asymptotics (Staiger and Stock, 1997), confirms that departures from normality depend strictly on three factors: strength of instruments, level of endogeneity and overidentification. Some of these problems could be exacerbated in the empirically–relevant just identified case, where the number of instruments is equal to the number endogenous variables. In fact, the inexistence of moments for IV estimator (Mariano, 2001) may produce extreme values for the structural parameters and huge standard errors.

1.3 Main contribution

The aim of this work is to analize the properties of bootstrap estimator in linear IV under weak instruments, both analitically and through Monte Carlo simulations, in order to study and capture sources of failure, verifying the informativeness of bootstrap in the context of weak instruments. We provide a new derivation of the bootstrap limiting distribution of the IV/TSLS estimator using two types of residual–based resampling schemes. Bootstrap estimators converges to a random limiting distribution, conditionally on the original data, preserving some components of the weak instrument asymptotics. Some examples are presented to show failures of bootstrap inference under weak instruments. In fact, bootstrap may provide information on the failures of standard inference caused by weak instruments, overidentification, high level of endogeneity and, in particular, a lack of moments of the estimators, that could be severely reflected in the bootstrap samples.

Subsequentely, a bootstrap normality test on a certain number of replications is introduced in the context of weakly identified linear models, adapting a framework introduced by Angelini et al. (2016). The main idea is to check failures in the standard regularity conditions, reasonably driven by weak instruments in IV setting. Well–known normality tests as could be applied and, in addition, a modified Shapiro-Wilk test with known mean (Hanusz et al., 2016) is proposed in order to control the distance from limiting distribution of the estimators, improving power of the tests in presence of more than one instrument. In fact, these estimators are asymptotically normal in overidentified models (Bekker, 1994) although consistency breaks down dramatically when the number of instruments is arbitrarily large. We also introduce standard normality tests based on the moments of standard gaussian random variable. The main idea is to detect issues in conventional standardization of the bootstrapped estimator, as suggested by Zhan (2017), also caused by the weak identification combined with the degree of overidentification and the endogeneity level. These kinds of large samples tests are based on the first four moments of the standard normal and have asymptotic chi square distributions under the null hypothesis. From a theoretical point of view they are similar to those proposed by Bontemps and Meddahi (2005).

This PhD Thesis is organized as follows. Chapter 2 presents the basic IV linear model with one endogenous regressor, weak instrument asymptotics and the bootstrap techniques for limited information system estimators. Chapter 3 examines analitically and through simulation study the bootstrapped distribution of IV/TSLS estimators, presenting a new derivation under weak instrument asymptotics. In Chapter 4 we firstly analyze some detection tools and then develop new bootstrap–based tests for weak instruments in linear IV models. These tests are analyzed through Monte Carlo simulations, in order to verify the performance of proposed methods, in terms of size and power of tests; at the end of this Chapter we apply these methods on two real cross–sectional data. Conclusions are in Chapter 5.

Chapter 2

Weak instruments and bootstrap in instrumental variable: a review

In this chapter we review the issue of weak instruments in instrumental variables linear models with one endogenous regressor, summarizing some available tests and procedures to detect failures in the relevance conditions, i.e. when the instruments are poorly correlated with the endogenous explanatory variable. This framework requires the so–called "weak instrument asymptotics", a particular nesting useful to study the asymptotic properties of estimators and test statistics. Furthermore, an extensive Monte Carlo simulation is conducted to highlight the performance of well–known estimators under low relevance, moderate overidentification and several degrees of endogeneity. Evidences regarding non–normality in commonly used estimators are discussed under different data generating processes (DGPs), considering skewed disturbances and non perfectly exogenous instruments. Asymptotic behaviour of associated t/Wald statistics under the null hypothesis is also presented under different DGPs. Simulation study includes an evaluation of first stage F test screening, in order to point out some issues in conventional procedures and rules of thumb applied by practitioners. We propose Kolmogorov–Smirnov (KS) distance as a further performance indicator in order to compare several IV estimators and test statistics against their non–normality, possibly generated by weak instruments.

Therefore, we illustrate bootstrap methods in instrumental variables linear models. In general resampling techniques could be useful to improve inference in point estimations, confidence intervals and test statistics. However, the bootstrap is not deemed always valid when instruments are collectively weak. Recent methods may be improve inference in some cases, especially in hypothesis testing, but relevance remains a crucial issue to discuss the usefulness of bootstrap inference in linear IV models.

2.1 Instrumental variables in Econometrics

Instrumental variables represent a very powerful tool in regression analysis when the explanatory variables are deemed correlated with the stochastic component. When endogeneity appears, Ordinary Least Squares (OLS) estimators give biased and inconsistent estimates of the parameters. Main sources of endogeneity refer to a) omitted variable bias, b) measurement error and c) reverse causality, also called simultaneous equations issue. IV estimators are widely used in Econometrics to overcome these kinds of problems in a broad range of empirical research. In microeconometrics, they are often applied in the context of wage equations, e.g. estimating the effects of education and experience on earnings. Several papers are developed in this field, including Grichiles (1977), Angrist and Kreueger (1991), Card (2001), Oreopoulos (2006) and Pischke and Von Wachter (2008). Some empirical works in economics regard the estimation of elasticity of intertemporal substitution (EIS), e.g. Yogo (2004) and Gomes and Paz (2011), using stationary time series data. Further studies are focused on political economics, including Acemoglu et al. (2008), or crime rates in U.S. (Levitt, 2002). In macroeconometrics, one of the first application concerns the estimation of Klein model (see for example Greene, 2000). Nowadays, the usage of IV estimation is still under debate in the context of linearized rational expectation models, especially in the estimation of New Keynesian Phillips Curve (NKPC) parameters (Kleibergen and Mavroeidis, 2009).

Nevertheless, IV estimation is strictly based on the sources of *valid* instruments. Validity essentially requires two properties: *exogeneity* and *relevance*; the first is usually verified through overidentification tests applying Sargan–type statistics or, alternatively, likelihood ratio. However, these procedures require exogeneity of a certain number of instruments, at least equal to the amount of endogenous regressors. Recently, some authors criticized this practice: Conley et al. (2012) approach this issue from a Bayesian point of view, while Ashley (2009) proposes sensitivity measures to evaluate inference under possibly endogenous instruments.

A large literature is mainly focused on the failure of relevance condition and essentially concerns two aspects. The first regards development of tests and detection methods for the weakness of instruments. The second is based on robust techniques, in order to obtain reliable point estimates, tests and confidence intervals under poorly relevant instruments. In this context, (partially) robust estimators are generally included in a broad group called $\kappa-class$, while robust tests are proposed by Kleibergen (2002) and Moreira (2003). Existing surveys on inference under weak instruments/weak identification and robust methods include Stock et al. (2002), Dufour (2003), Andrews and Stock (2005) and Poskitt et al. (2013).

2.1.1 The model

We consider the following linear model with two equations, containing one right–hand–side regressor (in the structural equation) and k instrumental variables, where k is considered as a fixed number. The system takes the following form:

$$\begin{aligned} y_t &= x_t\beta + u_t \\ x_t &= Z_t\boldsymbol{\pi} + v_t, \end{aligned} \tag{2.1}$$

where $t = (1, \ldots, T)$ and T is the sample size. Observations can be stacked to obtain:

$$\mathbf{y} = \mathbf{x}\beta + \mathbf{u} \tag{2.2}$$

$$\mathbf{x} = \mathbf{Z}\boldsymbol{\pi} + \boldsymbol{v}, \tag{2.3}$$

where (2.2) is the structural equation and (2.3) is called first stage equation. Model in (2.2, 2.3) includes an outcome variabile $\mathbf{y} = (y_1 \ldots, y_T)'$ while $\mathbf{x} = (x_1, \ldots, x_T)'$ is the endogenous explanatory variable, assuming $E(x_t u_t) \neq 0$, and $\mathbf{Z} = (Z_1, \ldots, Z_T)'$ is a full column rank matrix $T \times k$ containing exogenous (also called excluded) instruments. The scalar parameter of interest is β, while the $k \times 1$ vector $\boldsymbol{\pi}$ presents nuisance parameters associated to instrumental variables. Substituting (2.3) in (2.2), we obtain the so called reduced form as:

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\pi}\beta + \boldsymbol{\eta}, \tag{2.4}$$

where $\boldsymbol{\eta} = \mathbf{v}\beta + \mathbf{u}$.

The presence of a non–zero correlation between x_t and u_t makes ordinary least squares (OLS) estimator, defined as $\hat{\beta}^{OLS} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y}$, largerly biased in finite sample and (possibly) inconsistent, where the amount of bias depends on the degree of correlation between regressor and the structural error component. To summarize, instrumental variables have the main role to isolate the exogenous variation of the endogenous regressor. When the number of instruments is equal to one, the equation (2.3) is reduced to:

$$\mathbf{x} = \mathbf{z}\pi + \boldsymbol{v}, \tag{2.5}$$

where, in (2.5), π is reduced to a scalar parameter, $\mathbf{z} = (z_1, \dots, z_T)'$ represents the only instrumental variable and the reduced form becomes $\mathbf{y} = \beta\pi\mathbf{z} + \boldsymbol{\eta}$, where $\boldsymbol{\eta}$ is previously defined. Usually, if the number of IV is equal to the amount of endogenous regressors the model is called just–identified or perfectly identified.

Finally, we introduce some notation for the following paragraphs: the symbol "$\xrightarrow{p}$" denotes convergence in probability, while "$\xrightarrow{d}$" indicates convergence in distribution.

2.1.2 Assumptions

The vectors $\mathbf{u} = (u_1, \dots, u_T)'$ and $\boldsymbol{v} = (v_1, \dots, v_T)'$, which contain disturbances from both equations, are conditionally homoskedastic with zero mean and with the following 2×2 covariance matrix:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \cdot & \sigma_v^2 \end{pmatrix} = \begin{pmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \cdot & \sigma_v^2 \end{pmatrix}, \tag{2.6}$$

which is symmetric and positive definite. The scalar parameter $\rho = \text{Corr}\,(u_t, v_t)$ is the level of *endogeneity*, while the population variances of the disturbances are constant and finite, i.e. $\sigma_u^2 < \infty$ and $\sigma_v^2 < \infty$.

Two main assumptions regard *validity* of instruments, involving the exogeneity of Z_t (*exclusion* condition) and the so called *relevance* condition. Exogeneity implies $E(Z_t u_t) = E(Z_t v_t) = 0$, while relevance condition holds when $E(Z_t x_t) \neq 0$ and it is necessary for the identification of structural parameter β. In addition, the $k \times k$ covariance matrix of the instruments is a full rank matrix $E(Z_t Z_t') = \mathbf{Q}_{ZZ}$, and $\mathbf{Z}'\mathbf{Z}/T$ is a consistent estimator for $\mathbf{Q}_{ZZ}$ i.e. $\mathbf{Z}'\mathbf{Z}/T \xrightarrow{p} \mathbf{Q}_{ZZ}$. Furthermore, asymptotic normality of estimators requires $E(y_t^4) < \infty$, $(T^{-1}u_t'u_t, T^{-1}v_t'v_t) \xrightarrow{p} (\sigma_u^2, \sigma_v^2)$ and also $(T^{-1/2}Z_t'u_t, T^{-1/2}Z_t'v_t) \xrightarrow{d} (W_{Zu}, W_{Zv}) \sim N(\mathbf{0}, \boldsymbol{\Sigma} \otimes \boldsymbol{Q}_{ZZ})$, where "$\otimes$" denotes the Kronecker product. In the just identified case with a single instrument, $E(z_t z_t')$ is reduced to σ_z^2, $T^{-1}\mathbf{z}'\mathbf{z} \xrightarrow{p} \sigma_z^2$ and $(T^{-1/2}z_t'u_t, T^{-1/2}z_t'v_t) \xrightarrow{d} (w_{zu}, w_{zv}) \sim N(\mathbf{0}, \sigma_z^2\boldsymbol{\Sigma})$.

Finally, we point out that IV estimators are invariant for an orthonormal transformation of the instruments, such that $T^{-1}(\mathbf{Z}'\mathbf{Z}) = I_k$. We indicate with the subscript t both cross–section and time series data, while the latter also requires stationarity and ergodicity conditions of the variables.

Exogenous Covariates

System described in (2.2, 2.3) allows to include exogenous covariates, sometimes denoted as control variables or "included" instruments, in both equations. If there are l exogenous variable $\mathbf{W}$, where $\mathbf{W} = (W_1, \ldots, W_T)'$ is a $T \times l$ matrix and $Cov(W_t, u_t) = Cov(W_t, v_t) = 0$, the endogenous variables $(\mathbf{y}, \mathbf{x})$ and instruments may be substituted, in the estimation procedures, by residuals from their projection on $\mathbf{W}$. This could be done applying the Frisch-Waugh-Lovell (FWL) theorem (1963), as remarked in Stock et al. (2002). Thus, a more general model could be expressed as follows:

$$\begin{aligned} \mathbf{y} &= \mathbf{x}\beta + \mathbf{W}\mathbf{\Gamma}_1 + \boldsymbol{u} & (2.7) \\ \mathbf{x} &= \mathbf{Z}\boldsymbol{\pi} + \mathbf{W}\mathbf{\Gamma}_2 + \boldsymbol{v}, & (2.8) \end{aligned}$$

where Γ_1 and Γ_2 are vectors $l \times 1$ containing parameters associated to control variables. Equations in (2.8, 2.7) could be rewritten in the following way:

$$\begin{aligned} \tilde{\mathbf{y}} &= \tilde{\mathbf{x}}\beta + \tilde{\boldsymbol{u}} \\ \tilde{\mathbf{x}} &= \tilde{\mathbf{Z}}\boldsymbol{\pi} + \tilde{\boldsymbol{v}}, \end{aligned}$$

where $\tilde{\mathbf{x}} = M_W\mathbf{x}$, $\tilde{\mathbf{y}} = M_W\mathbf{y}$, $\tilde{\mathbf{Z}} = M_W\mathbf{Z}$, $(\tilde{\mathbf{u}}, \tilde{\mathbf{v}}) = (M_W\mathbf{u}, M_W\mathbf{v})$ and $M_W = (I_l - P_W)$ is a symmetric and idempotent matrix.

2.1.3 Estimation

Considering system in (2.2, 2.3), a common choice in the estimation of (scalar parameter) β is two–stage least squares (TSLS) estimator, defined as follows:

$$\begin{aligned} \hat{\beta}_T^{TSLS} &= \left[\mathbf{x}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{x}\right]^{-1}\left[\mathbf{x}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}\right] \\ &= \left(\mathbf{x}'P_Z\mathbf{x}\right)^{-1}\left(\mathbf{x}'P_Z\mathbf{y}\right), & (2.9) \end{aligned}$$

where $P_Z = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'$ is the symmetric and idempotent projection matrix of instruments. Vector of nuisance parameters $\boldsymbol{\pi}$ could be estimated through OLS:

$\hat{\pi}_T^{OLS} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{x}$. Considering $\boldsymbol{\pi} = \boldsymbol{\pi}_0 \neq 0$, where $\boldsymbol{\pi}_0$ is fixed, and if the assumptions of section 2.1.2 hold, TSLS in (2.9) is consistent and asymptotically normal distributed. This type of modeling is known in IV literature as *strong instrument* asymptotics. In particular: $\hat{\beta}_T^{TSLS} - \beta \xrightarrow{p} 0$ with the following asymptotic distribution:

$$T^{1/2}(\hat{\beta}_T^{TSLS} - \beta) \xrightarrow{d} N(0, \sigma_u^2 [\mathbf{x}' P_Z \mathbf{x}]^{-1}), \tag{2.10}$$

where $\sigma_u^2 = E(u_t u_t')$ and the estimation error of $\hat{\beta}_T^{TSLS}$ as $\hat{\beta}^{TSLS} - \beta = (\mathbf{x}' P_Z \mathbf{x})^{-1} (\mathbf{x}' P_Z \mathbf{u})$. Recalling $\boldsymbol{\pi} = \boldsymbol{\pi}_0$ and $\mathbf{Z}'\mathbf{Z}/T \xrightarrow{p} \mathbf{Q}_{ZZ}$, the expression in (2.10) may be modified in the following way:

$$T^{1/2}(\hat{\beta}_T^{TSLS} - \beta) \xrightarrow{d} N(0, \sigma_u^2 [\boldsymbol{\pi}_0' \mathbf{Q}_{ZZ} \boldsymbol{\pi}_0]^{-1}). \tag{2.11}$$

In the empirically–relevant just identified case, TSLS is reduced to the so-called instrumental variable estimator (IV):

$$\hat{\beta}_T^{IV} = (\mathbf{z}'\mathbf{x})^{-1} \mathbf{z}'\mathbf{y} = \frac{\sum_{t=1}^T z_t y_t}{\sum_{t=1}^T z_t x_t}, \tag{2.12}$$

and its asymptotic variance could be estimated using the following quantity $\hat{\omega}$:

$$\hat{\omega} = \hat{\sigma}_u^2 \frac{\sum_{t=1}^T z_t^2}{\sum_{t=1}^T (x_t z_t)^2}.$$

Under strong instruments, i.e. $\pi = \pi_0$, estimator in (2.12) is consistent, and its asymptotic distribution becomes:

$$T^{1/2}(\hat{\beta}_T^{IV} - \beta) \xrightarrow{d} N\left(0, \omega^2\right). \tag{2.13}$$

where the numerator of asymptotic variance is $\omega^2 = \sigma_u^2(\sigma_z^2 \pi_0^2)^{-1}$. and $\hat{\omega}^2 \xrightarrow{p} \omega^2$

Recalling that $\hat{\mathbf{x}} = P_Z\mathbf{x}$, it is noticeable that TSLS estimator in (2.9) could be rewritten as:

$$\hat{\beta}_T^{TSLS} = (\hat{\mathbf{x}}'\mathbf{x})^{-1}\hat{\mathbf{x}}'\mathbf{y}. \tag{2.14}$$

Following (2.14), TSLS can be viewed as an IV estimator where the endogenous regressor $\mathbf{x}$ is instrumented with its fitted values $\hat{\mathbf{x}} = P_Z\mathbf{x}$. This estimator could be also written as $\hat{\beta}_T^{TSLS} = (\hat{\mathbf{x}}'\hat{\mathbf{x}})^{-1}\hat{\mathbf{x}}'\mathbf{y}$, where $\hat{\mathbf{x}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}\mathbf{x}$, (see Wooldridge, 2010 for details). The denomination Two Stage Least Squares comes directly from this equality.

When the system allows to include exogenous regressors, the estimator in (2.9) becomes:

$$\begin{aligned}\hat{\beta}_T^{TSLS} &= \left[\tilde{\mathbf{x}}'\tilde{\mathbf{Z}}(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-1}\mathbf{Z}'\tilde{\mathbf{x}}\right]^{-1}\left[\tilde{\mathbf{x}}'\tilde{\mathbf{Z}}(\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}'\tilde{\mathbf{y}}\right] \\ &= \left(\tilde{\mathbf{x}}'P_{\tilde{Z}}\tilde{\mathbf{x}}\right)^{-1}\left(\tilde{\mathbf{x}}'P_{\tilde{Z}}\tilde{\mathbf{y}}\right),\end{aligned}$$

and IV in (2.12) takes the following form:

$$\hat{\beta}_T^{IV} = \left(\tilde{\mathbf{z}}'\tilde{\mathbf{x}}\right)^{-1}\tilde{\mathbf{z}}'\tilde{\mathbf{y}} = \frac{\sum_{t=1}^{T}\tilde{z}_t\tilde{y}_t}{\sum_{t=1}^{T}\tilde{z}_t\tilde{x}_t},$$

where the quantities denoted with the tilde are previously defined in terms of residual projections of the included exogenous instruments $\mathbf{W}$. Under strong instrument asymptotics both estimators are consistent and asymptotically normal distributed.

2.1.4 A wide class of estimators

The previously discussed TSLS and IV estimator, defined in (2.9) and (2.12), are nested in a wider group, called $\kappa-$class, introduced by Nagar (1959). Considering the system in (2.2, 2.3) with k instruments, a generic $\kappa-$class estimator can be written as follows:

$$\begin{aligned}\hat{\beta}_T^{\kappa-\text{cl}} &= \left[\mathbf{x}'(I-\kappa M_Z)^{-1}\mathbf{x}\right]^{-1}\mathbf{x}'(I-\kappa M_Z)^{-1}\mathbf{y}, \\ &= (\mathbf{A}\mathbf{x})^{-1}\mathbf{A}\mathbf{y} \qquad (2.15)\end{aligned}$$

where $\mathbf{A} = \mathbf{x}'(I_k - \kappa M_Z)^{-1}$ is a matrix expressed in terms of $\mathbf{x}$ and $\mathbf{Z}$, and also $M_Z = (I_k - P_Z)$. From expression (2.15), it is noticeable that TSLS and IV are special cases of $\kappa-$class estimators where κ is fixed and equal to 1, and then $\mathbf{A} = \mathbf{x}'P_Z$. OLS estimator $\hat{\beta}_T^{OLS} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y}$ is also nested in this class, having $\kappa = 0$ and $\mathbf{A} = \mathbf{x}'$. An estimator presenting $\hat{\kappa} = T/(T-k+2)$ is called Bias adjusted TSLS (BTSLS). It is mainly used in presence of more instruments and its properties are analized and discussed in Donald and Newey (2001).

The so–called limited information maximum likelihood (LIML) is another $\kappa-$class estimator where $\kappa = \hat{\kappa}_{LIML}$ is the smallest eigenvalues of the matrix: $(\mathbf{x}'\mathbf{x} - \kappa\cdot\mathbf{x}'M_Z\mathbf{x})$. Performance of LIML estimator are analyzed through simulation by Hahn and Inoue (2002). A slightly modification of the LIML was introduced by Fuller (1977) using

$\kappa_{Full} = \hat{\kappa}_{LIML} - \underline{c}/(T-k)$, where $\underline{c}$ is a fixed constant chosen by the researcher, often setted equal to an integer number (1 or 4). In this context it is important to point out that LIML and Fuller share the same asymptotic variance of IV/TSLS estimator and are also asymptotically normal distributed under strong instrument asymptotics, i.e. $\boldsymbol{\pi} = \boldsymbol{\pi}_0 \neq 0$. Therefore, some jackknife–based estimators are nested in this class; we remind JIVE proposed by Angrist et al. (1999). Furthermore in the just–identified case LIML, TSLS and BTSLS estimators coincide with IV.

LIML, BTSLS or Fuller can be applied when the number of instruments is moderately large, and their behaviour in finite samples is analized and discussed in some papers including Blomquist and Dahlberg (1999) and Hahn et al. (2004). These works also mention issues regarding the absence of finite moments in $k--$class, occurring especially in LIML and BTSLS. We do not discuss theoretically this problem, but our empirical evidences show that it could dramatically affect the distribution of estimators and tests in finite samples, if combined with weak or irrelevant instruments.

More than one endogenous regressor

When the number of endogenous explanatory variable m is greater than 1, the system described in (2.2, 2.3) contains $(m+1)$ equations and takes the following form:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{X} &= \mathbf{Z}\boldsymbol{\Pi} + \boldsymbol{V}, \end{aligned} \tag{2.16}$$

where $\boldsymbol{\Pi}$ is a matrix $k \times m$ of nuisance parameters, $\boldsymbol{\beta}$ is a vector $m \times 1$ of structural parameters, $\mathbf{X}$ is a matrix $T \times m$ and $\mathbf{Z}$ is again the $T \times k$ matrix of instruments. Given the system in (2.16), vector $\boldsymbol{\beta}$ is identified if $k \geq m$ and TSLS estimator, previously defined in the (2.9), becomes:

$$\hat{\beta}_T^{TSLS} = (\mathbf{X}'P_Z\mathbf{X})^{-1}\mathbf{X}'P_Z\mathbf{y} = (\hat{\mathbf{X}}'\mathbf{X})^{-1}\hat{\mathbf{X}}'\mathbf{y}. \tag{2.17}$$

In the just identified case, i.e. when the number of instruments is equal to the endogenous regressors ($k = m$), the (2.17) is reduced to IV estimator: $\hat{\boldsymbol{\beta}}_T^{IV} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}$. Assumptions regarding consistency and asymptotic normality are similar to those previously introduced in section 2.1.2, where now $\boldsymbol{V} = (V_1, \ldots, V_T)'$ is a matrix $T \times k$ and the full rank $(m+1) \times (m+1)$ covariance matrix of (conditionally homoskedastic) stochastic components is:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_u^2 & \Sigma_{Vu} \\ \Sigma_{uV} & \Sigma_{VV} \end{pmatrix}.$$

A simple example of linear two–endogenous variable system ($m = 2$), without exogenous covariates, is the following:

$$\begin{aligned} \mathbf{y} &= \mathbf{x}_1\beta_1 + \mathbf{x}_2\beta_2 + \mathbf{u} \\ \mathbf{x}_1 &= \mathbf{Z}\boldsymbol{\pi}_1 + \boldsymbol{v}_1 \\ \mathbf{x}_2 &= \mathbf{Z}\boldsymbol{\pi}_2 + \boldsymbol{v}_2. \end{aligned}$$

and the variance–covariance matrix of disturbances is the 3×3 matrix:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_u^2 & \sigma_{v_1u} & \sigma_{v_2u} \\ \sigma_{v_1u} & \sigma_{v_1}^2 & \sigma_{v_1v_2} \\ \sigma_{v_2u} & \sigma_{v_1v_2} & \sigma_{v_2}^2 \end{pmatrix}.$$

Finally we remark that all $\kappa-$class estimators, defined in the (2.15), may be rewritten in the multiple endogenous variable case, and takes the following form:

$$\begin{aligned} \hat{\beta}_T^{\kappa-\mathrm{cl}} &= \left[\mathbf{X}'(I - \kappa M_Z)^{-1}\mathbf{X}\right]^{-1}\mathbf{X}'(I - \kappa M_Z)^{-1}\mathbf{y}, \\ &= (\mathbf{AX})^{-1}\mathbf{Ay}. \end{aligned}$$

2.2 Weak instruments

When the instruments are deemed weak the empirical distribution of the TSLS/IV estimators may be very far from the limiting distribution. Failure of conventional asymptotics is also reflected in hypothesis testing, as pointed out by Nelson and Startz (1990a), and confidence intervals (Zivot et al., 1998) resulting too wide. In this section we briefly introduce weakly identified models in econometrics by using the sequence of modeling denoted as "weak instrument asymptotics". Then, we discuss detection methods and tests for weak instruments in linear IV models and briefly mention some advances in this topic. Pratictioners mainly apply procedures based on the first stage F statistic, testing the null hypothesis that all the coefficients in vector $\boldsymbol{\pi}$ are equal to zero.

2.2.1 Weakly identified models

Weak identification is not a prerogative of linear IV, arising in many econometric models. Some examples include AutoRegressive Moving Average (ARMA) with near canceling unit root, Structural Vector AutoRegressions (SVARs), rational expectation models and linear or non–linear GMM, including IV/TSLS inference as a particular case. In the latter case, the source of this problem is related to a lack of information to estimate the parameters of interest. In the IV context Andrews and Mikusheva (2014) show that this issue clearly appears in estimation via maximum likelihood, in which the occurring disparity between estimated information measures, i.e. the expressions for variance of score statistic, could be interpreted as signal of weak identification. In case of non–linear GMM the interpretation may be different; briefly speaking it concern the randomness of curvature in objective (criterion) function. From a different point of view, Nelson and Startz (2007) introduce the Zero Information Limit Conditions (ZILC) to show the cases whose weak identification lead to spurious inference, as occurs especially in IV where conventional estimators presents unreliable precision.

In mostly applied liner model with one endogenous regressor, weak identification practically means that the excluded instruments are not collectively correlated with the troublesome explanatory variable, i.e. $E(Z_t x_t) \approx 0$. To summarize, the instrumental variables are not useful to predict the endogenous explanatory variable. Then, first stage R^2, denoted as R_f^2 may be very close to zero in finite samples. Under these scenarios standard inference provides poor first-order approximation, even in very large samples, as highlighted by Bound et al. (1995) reviewing the estimates from seminal paper of Angrist and Kreueger (1991).

Unidentification and intuition

In special unidentified case, when $E(Z_t x_t) = 0$ and $\boldsymbol{\pi} = 0$, scalar parameter β is completely not identified and the asymptotic distribution of TSLS/IV estimator could be non–normal, even lacking any finite moment. In order to give a first intuition of unidentfication, we consider the case of a single irrelevant instrument, i.e. $\pi = 0$ and then $\mathbf{x} = \boldsymbol{v}$. First of all, IV estimator defined in (2.12) is distributed as a ratio of two correlated gaussian (Cauchy–like random variable), as follows:

$$\hat{\beta}_T^{IV} - \beta = \frac{\sum_{t=1}^{T} z_t u_t}{\sum_{t=1}^{T} z_t v_t} \xrightarrow{d} \frac{w_{zu}}{w_{zv}}. \tag{2.18}$$

where $(w_{zu}, w_{zv})' \sim N(0, \sigma_z^2 \mathbf{\Sigma})$ and $\mathbf{\Sigma}$ is defined in section 2.1.2. In this extreme case IV estimator is not consistent because does not converge in probability to the true value. Therefore, considering $\sigma_v^2 = \sigma_u^2 = \sigma_z^2$ and defining $w = w_{zu} - \rho w_{zv}$, where $w \sim N(0, (1-\rho^2))$, we obtain the following result:

$$\hat{\beta}_T^{IV} - \beta \xrightarrow{d} \rho + \frac{w}{w_{zv}}. \tag{2.19}$$

By expression (2.19), we notice that ratio w/w_{zv} is now a proper Cauchy random variable; any inference conducted through usual procedures may easily fail.

Weak instruments problem arises when the relevance condition is close to a failure rather than being completely violated, as illustrated in the previous case. To show this fact, we consider again the just–identified case where the relevance condition is not violated, but the correlation between instrument and exogenous regressor may be very small. In this situation, since the IV estimator in (2.12) can be straightforward expressed as $\hat{\beta}_T^{IV} = Cov(z_t, y_t)/Cov(x_t, z_t)$, its estimation error can be re-written as the following ratio of covariances:

$$\hat{\beta}_T^{IV} - \beta = \frac{\sum_t y_t z_t}{\sum_t x_t z_t} = \frac{Cov(z_t, u_t)}{Cov(z_t, x_t)}. \tag{2.20}$$

Estimation error in (2.20) presents huge values if the covariance (correlation) between instrument and regressor vanishes to zero.

From another point of view, substituting x_t in the second stage, i.e. first equation of (2.12), we obtain $y_t = \pi\beta x_t + \eta_t$. Denoting $\pi\beta = \delta$ as the so called reduced–form (recalling expression (2.4)) parameter, then $\beta = \delta/\pi$ and weak identification issue clearly appears if $\pi \approx 0$. Finally, in the overidentified case with $k > 1$, weak instruments practically means that $rank[E(\mathbf{Z}'\mathbf{x})] \approx 0$.

2.2.2 Weak instrument asymptotics

Weak instrument asymptotics is a sequence of models introduced by Staiger and Stock (1997) and developed by Stock and Yogo (2005) in order to study the properties of IV/TSLS estimators under weak instruments. The basic idea is to set the vector of nuisance parameters in a $1/\sqrt{T}$ neighbourhood of 0: $\boldsymbol{\pi} = \boldsymbol{\pi}_T = C/\sqrt{T}$, where C is $k \times 1$ vector of constants. Thus, under this nesting, $\boldsymbol{\pi}$ is drifting to zero as $T \to \infty$.

Adopting weak instrument asymptotics the system in (2.2, 2.3) becomes:

$$\begin{aligned} \mathbf{y} &= \mathbf{x}\beta + \mathbf{u} && (2.21)\\ \mathbf{x} &= \mathbf{Z}\frac{C}{\sqrt{T}} + \boldsymbol{v}. && (2.22) \end{aligned}$$

Starting from system (2.21, 2.22) and given the assumptios of Section 2.1.2, Staiger and Stock (1997) derived the distribution of TSLS under weak instrument asymptotics as follows:

$$\begin{aligned} (\hat{\beta}_T^{TSLS} - \beta) &= (\mathbf{x}P_Z\mathbf{x})^{-1}\mathbf{x}P_Z\mathbf{u}\\ &\xrightarrow{d} \frac{(\mathbf{Q}_{ZZ}C + W_{Zu})'\mathbf{Q}_{ZZ}^{-1}W_{Zv}}{(\mathbf{Q}_{ZZ}C + W_{Zv})'\mathbf{Q}_{ZZ}^{-1}(\mathbf{Q}_{ZZ} + W_{Zv})}\\ &= \frac{(\lambda + W_{Zu})'W_{Zv}}{(\lambda + W_{Zv})'(\lambda + W_{Zv})}; \qquad (2.23) \end{aligned}$$

where $(W_{Zu}, W_{Zv}) \sim N(0, \mathbf{\Sigma})$ and $\lambda = C'\mathbf{Q}_{ZZ}^{1/2}$. By using $\boldsymbol{\nu}_1 = (\lambda + W_{Zu})'$ and $\boldsymbol{\nu}_2 = (\lambda + W_{Zv})'(\lambda + W_{Zv})$, the expression in 2.23 modifies to:

$$(\hat{\beta}_T^{TSLS} - \beta) \xrightarrow{d} \boldsymbol{\nu}_1^{-1}\boldsymbol{\nu}_2.$$

In weak instruments setting constant C in (2.23) is often substituted by $\sqrt{\mu^2}$, where μ^2 represents the so–called population *concentration* parameter, introduced by Rothenberg (1984). He derive an alternative representation of TSLS estimator error under fixed instruments and gaussian disturbances, showing that it strictly depends on the quantity $\boldsymbol{\pi}'\mathbf{Z}'\mathbf{Z}\boldsymbol{\pi}$. Concentration parameter plays a central role in the relevance of instruments and it is defined as follows:

$$\mu^2 = \frac{\boldsymbol{\pi}'\mathbf{Z}'\mathbf{Z}\boldsymbol{\pi}}{\sigma_v^2}. \qquad (2.24)$$

Weak instrument asymptotics based on $C = \mu$ implies that μ^2, representing the strength of instruments, tends to a constant even if sample size diverges, i.e $T \to \infty$. Relevance of all instruments is measured by the average population concentration parameter μ^2/k, where $k > 1$ is a fixed number of excluded instruments. In the limit case of irrelevance, i.e. $\mu^2/k = 0$, structural coefficient β is not point identified, as previously discussed, while the relevance level of the instruments grows as $\mu^2/k \to \infty$, reducing to strong instrument asymptotics.

In the just identified case, when $\mathbf{Q}_{ZZ} = \sigma_z^2$, the interpretation of weak instrument asymptotics based on $\pi = \pi(T) = cT^{-1/2}$ is even more immediate. Recalling the denominator of IV estimator in (2.12), it could be expressed as follows:

$$\frac{1}{T}\sum_{t=1}^{T} x_t z_t = \frac{c}{T^{1/2}}\frac{1}{T}\sum_{t=1}^{T} z_t^2 + \frac{1}{T}\sum_{t=1}^{T} v_t z_t, \tag{2.25}$$

and then expression in (2.25) becomes an $O_p(T^{-1/2})$ rather than being $O_p(1)$. Then, IV estimator under weak instrument asymptotics is:

$$\hat{\beta}_T^{IV} - \beta \xrightarrow{d} \frac{w_{zu}}{(c\sigma_z^2 + w_{zv})}, \tag{2.26}$$

where $(w_{zu}, w_{zv}) \sim N(0, \sigma^2 \mathbf{Q}_{ZZ})$. Finally, the expression for concentration parameter in just identified case takes the following expression:

$$\mu^2 = \frac{\pi^2 \sigma_z^2 T}{\sigma_v^2}, \tag{2.27}$$

and the simplified case of $\sigma_z^2 = \sigma_v^2 = 1$, then μ^2 is reduced to $\pi^2 T$.

2.2.3 Tests and detection methods

Since the seminal paper of Angrist and Kreueger (1991), practitioners face the problem of detecting poorly relevant instruments. First of all, Bound et al. (1995) argue that both first stage F statistic and first stage R^2, denoted as R_f^2, should be used to diagnose complete failures in relevance of instruments. Since R_f^2 could be a controversial measure in presence of control variables, several works propose *partial* R_f^2, introduced by Shea (1997), in order to isolate the explanatory power of instruments. However, this practice is still under debate because R^2 measures strictly depend on the sample size.

From a different point of view, Hahn and Hausman (2002a) propose a specification test in IV framework, based on the so called reverse estimator of TSLS: $\hat{\beta}_T^{REV} = (\mathbf{y}'P_Z\mathbf{y})^{-1}\mathbf{y}'P_Z\mathbf{x}$, using the asymptotically equivalence, occurring under strong instruments, between $\hat{\beta}_T^{TSLS}$ and $1/(\hat{\beta}_T^{REV})$. Despite existing methods this test presents the null hypothesis of strong instruments but has very low power under poor relevance, as illustrated by Hausman et al. (2005).

First stage F

The first stage F statistic on the excluded instrumental variable is typically used to assess the weakness of instruments, and it is defined as follows:

$$F = \frac{\hat{\pi}_T' \mathbf{Z}'\mathbf{Z}\hat{\pi}_T / k}{\hat{\sigma}_v^2}. \tag{2.28}$$

If the first stage F test rejects the null hypothesis of $\boldsymbol{\pi} = 0$, there is an empirical evidence that relevance condition holds. However, this does not guarantee that instruments are not weak, as pointed out in simulation study by Hall et al. (1996). Moreover, first stage F represents a crucial quantity in the estimation of (unknown) population μ^2. In fact, considering overidentified models, the average concentration parameter can be estimated using the following relationship:

$$\hat{\mu}/k = (F - 1). \tag{2.29}$$

The result in (2.29) comes from the approximate expected values of the F statistic. As Staiger and Stock (1997) pointed out, $E(F) \cong 1 + \mu^2/k$, and $F - 1$ can be viewed as an estimator of the average strength of the instruments.

Under weak instrument asymptotics F is asymptotically distributed as a noncentral $\chi_k^2(\mu^2/k)$, where k is the number of excluded instruments and noncentrality parameter is equal to μ^2/k. Stock and Yogo (2005) introduce quantitative definition of weak instruments considering both the relative bias of IV estimators, with respect to OLS, and relative size of the associated t/Wald test for the null hypothesis $H_0 : \beta = \beta_0$. They also compute critical values for the non central χ^2 distributions through Monte Carlo simulation, considering different combinations between instruments and endogenous regressors. This method has recently been improved analitically by Skeels and Windmeijer (2016).

To summarize, critical values for the F test are all approximately around ten. This generates a useful rule of thumb for economists and practitioners: instruments are deemed collectively strong or relevant if $F > 10$, and weak otherwise. In order to understand the intuition of this procedure we introduce the definition of bias (and relative bias) of TSLS estimator. Considering $\sigma_{uv} = \rho$ and $\sigma_v = 1$, te bias of TSLS quantity can be approximated (Hahn and Hausman, 2002b) by the following expression:

$$E(\beta_T^{TSLS}) - \beta \approx \rho \frac{k-2}{\mu^2}. \tag{2.30}$$

By the (2.30), approximate TSLS bias depends on the number of instruments k, concentration parameter μ^2 and degree of endogeneity. Hence, the procedure of Stock and Yogo (2005) is based on the *relative bias* between TSLS and OLS, approximable as follows:

$$\text{RelBias} \approx \frac{(k-2)(\mu^2 T^{-1} + 1)}{\mu^2},$$

which depends only on μ^2 and the sample size T, because the level of endogeneity affects both estimators in their finite sample properties. We further present results of our simulation study in order to show discrepancy between the estimated bias of TSLS and its approximated population counterpart, occurring if instruments are irrelevant ($\mu^2 = 0$) or very weak. If there are multiple endogenous regressors, this procedure could be generalized considering the multivariate counterpart of μ^2, called concentration matrix:

$$\boldsymbol{\mu^2} = T(\boldsymbol{\Sigma}_{VV}^{-1/2} \boldsymbol{\Pi}' \mathbf{Q}_{ZZ} \boldsymbol{\Pi} \boldsymbol{\Sigma}_{VV}^{-1/2}).$$

In this case, a statistic introduced by Cragg and Donald (1993) is used, representing the multivariate version of the First stage F.

However, these methods are based on strictly assumptions, as i.i.d. normal errors and fixed instruments, and could be sensitive to non-standard conditions, e.g. clustered robust errors and autocorrelation. Since heteroskedasticity in IV estimation is still of practical interest variable estimation (Hausman et al., 2012), the detection of weak instruments through a robust F statistic is still under debate. Monte Carlo results in Bun and de Haan (2010) suggest that $F > 10$ could be considered a low benchmark in presence of grouped errors, because F shows high values when the number of groups increases, while the clustered robust F decreases, as expected, but mainly underestimates the (true) concentration parameter μ^2. Similar results could be seen if the disturbances are simulated through autoregressive processes. Moreover, as pointed out by Hahn and Hausman (2003) and highlighted in our further simulations, F statistic does not take account of the endogeneity level, even if high values of ρ clearly affects the finite sample distribution of the estimators (and associated test statistics) in terms of bias as we see in expression (2.30). This arises even in case of homoskedastic disturbances (and instruments).

Further developments in First stage F

Extensions of first stage F methods are essentially developed in two areas: the former regards inference under heteroskedasticity (or autocorrelation) of the stochastic component, the latter is related to F test in case of multiple endogenous regressors. To overcome limitation of conventional screening, based on i.i.d. normal errors, some tests and procedures have been proposed. Primarily,Kleibergen and Paap (2006) develop a robust version of the Cragg–Donald statistic for unidentification in the case of more endogenous regressors ($m > 1$). Recently, Olea and Pflueger (2013) criticize the practice of comparing robust F statistic with conventional critical values. They propose a non–robust F with no–integer degrees of freedom, only valid in case of one endogenous regressor.

An F statistic in presence of multiple endogenous regressors is suggested by Angrist and Pischke (2008): they propose a *conditional* F in order to evaluate the strength of identification on each structural coefficient. The procedure essentially consists in replacing $(m - 1)$ endogenous variables with their reduced–form predictions, leading back to the univariate case. Despite usefulness of this new F test, replacing partial R_f^2 in well–known econometric software, they do not derive asymptotic theory and proper critical values. Moreover, Sanderson and Windmeijer (2016) introduce a new *conditional* F, correcting the proposal from Angrist and Pischke to have a proper asymptotic distribution under the null hypothesis.

2.3 Finite samples evidences under weak instruments

In this section an extensive Monte Carlo simulation is conducted in order to show the behaviour of conventional inference under weak instrument asymptotics. The simulation study has the main purpose to highlight the performance of IV/TSLS estimators under different strength of instruments combined with several endogeneity levels. The evaluation is conducted through some indicators recommended in the literature, to see in which scenarios the distribution of estimators deviates from their limiting distributions. In this context Kolmogorov–Smirnov statistic, based on the maximal difference between CDFs, is proposed as a further goodness of fit indicator. We also consider other $\kappa-$class estimators including LIML and two types of Fuller with $\underline{c} = 1, 4$, summarized in Section 2.1.4. Furthermore, non–normality of t–statistic under the null hypothesis, generated by inconsistency of estimators, is presented through simulation and graphic evaluation. Some evidences from the first stage tests and screening methods are highlighted, especially considering high endogeneity and non–gaussian disturbances.

First of all, we consider the just–identified case of one instruments–one regressor; successively we generate models with more than one instrument, combining weak identification with non–standard conditions, as non–normally distributed Z_t or (skewed) disturbances, and accounting a slightly violation of *exogeneity* condition, defined in Section 2.1.2.

2.3.1 Performance of estimators

The performances of IV/TSLS estimators under different degrees of identification and (positive) increasing endogeneity levels are evaluated through the following measures: Median point estimates (Median), Median Absolute Deviation (MAD), Root Mean Square Error (RMSE), Coverage rates (95%), Interdecile Range (IDR) and Kolmogorov Smirnov distance (KS). Some of these measures are also recommended in the literature, e.g. in works of Hahn et al. (2004) and Flores-Lagunes (2007). In particular, median point is reported to quantify distance from true value β, while MAD is considered a more robust measure of variability than standard error, especially in IV case because it present lack of all moments generating *extreme* values in the simulation.

We also investigate non–normality of the standardized distribution of the κ class: we present sampled moments of the Monte Carlo distribution of estimators, i.e. $\left\{\tilde{\beta}_{T,m}\right\}_{m=1}^{M}$, as mean, median, variance, kurtosis and skewness. InterQuartile Range (IQR) and KS distance are also reported.

Kolmogorov–Smirnov distance

In our simulation study KS distance is applied in order to measure the distance, in finite samples, between empirical cumulative distribution function (Ecdf) of TSLS/IV estimators and the gaussian Cdf. We follow the approach of Zhan (2017), computing KS in a bootstrap perspective, and extend to the $\kappa-$class estimators. KS statistic could be written as follow:

$$KS = \sup_{-\infty < c < +\infty} \left| P\left(\tilde{\beta}_T \leq c\right) - \Phi(c) \right|, \tag{2.31}$$

where $\tilde{\beta}_T$ is the so called "non–studentized" statistic:

$$\tilde{\beta}_T = \sqrt{T}\omega^{-1}\left(\hat{\beta}_T - \beta\right). \tag{2.32}$$

The quantity in (2.32) is standardized using the (true) value of asymptotic variance $\omega^2 = \sigma_u^2(\boldsymbol{\pi}\mathbf{Q}_{ZZ}\boldsymbol{\pi})^{-1}$, $\Phi(.)$ is the Cdf of standard normal distribution and $KS \in (0,1)$.

From a practical perspective, KS represents a measure of worst-case size distortion of the standardized statistic using the critical values from the normal distribution. First of all, KS increases with the level of endogeneity and the number of instruments, which are desirable features in the evaluation of IV/TSLS bias. Secondly, for given values of ρ and k, $KS \to 0$, where the strength of instruments diverges i.e. $\mu^2 \to \infty$, representing the limit case of strong instrument asymptotics. Low values of KS suggest the usefulness of conventional TSLS/IV or $\kappa-$class inference, whereas, when KS increases there is an evidence of sensitivity to weak instruments. Normality of t–statistics associated to IV/TSLS and partially robust estimators may be also evaluated in the same way. Finally this approach could be used also applied to compare performance of overidentifying restriction tests under weak instruments.

2.3.2 Set up of Monte Carlo

The main design presents a single instrument, drawn from a standard normal i.e. $z_t \sim N(0,1)$, and disturbances (u_t, v_t) sampled from a bivariate normal $N_2(\mathbf{0}, \boldsymbol{\Sigma})$ with the following covariance matrix:

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho \\ \cdot & 1 \end{pmatrix}.$$

The number of replicated datasets is equal to $M = 100000$ and observations are generated in the following way:

$$\begin{cases} y_t = \beta \cdot x_t + u_t \\ x_t = z_t(\mu T^{-1/2}) + v_t, \end{cases}$$

where $\pi = \mu T^{-1/2}$ and the scalar parameter of interest β is set to 0. Three sample sizes are considered: $T = \{100, 250, 1000\}$ and estimation is conducted through IV estimator, previously defined in (2.12). Positive endogeneity, expressed by ρ, varies from low to high: $\rho \in \{0.25, 0.5, 0.75, 0.9\}$. Concentration parameter μ^2, defined in the (2.24), takes following values: $\mu^2 \in \{0, 1, 5, 10, 20, 40, 60, 100\}$, where $\mu^2 = 0$ corresponds to a completely irrelevant instrument, and $\mu^2 = 100$ represents strong correlation between z_t and x_t. Therefore, different values of the scalar parameter $\pi = \pi(T) = \mu T^{-1/2}$ are generated under weak instrument asymptotics. These values imply standard errors of the IV estimator, becoming: $\sqrt{\sigma_u^2/(\mu^2 T^{-1})}$. Clearly, this quantity does not exist in the unidentified case of $\pi = 0$.

When instruments are more than one, the relevance is measured by the averaged concentration parameter μ^2/k. In this case the elements of the vector $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_k)'$ are modeled in the following way:

$$\pi_j = \sqrt{\frac{\mu^2}{k \cdot T}}, \quad \text{where} \quad j = \{1, \ldots, k\}\,.$$

An equivalent strategy consists in setting $\pi_1 = \sqrt{\mu^2}$ and then $\pi_j = 0$ for $j = 2, \ldots, k$. The number of considered instruments is $k = \{3, 5\}$, corresponding to a moderate degree of overidentification, in order to avoid the so called many instruments problem (Bekker, 1994). The Z_t's are drawn from a multivariate normal distribution $N_k(\mathbf{0}, I_k)$ where I_k is an identity $k \times k$ matrix and $k = 3, 5$. Then, the second equation of the system becomes $x_t = Z_t \boldsymbol{\pi} + v_t$. Strength of instruments μ^2/k takes values in $\{0, 1, 5, 10, 20, 40, 60, 100\}$, where again $\mu^2/k = 0$ represents a situation of totally irrelevant instruments.

We further consider LIML and Fuller estimators, introduced in subsection 2.1.4, to analyze their performance in finite samples under weak instruments asymptotics and overidentification. The constant $\underline{c}$ for Fuller estimator is set equal to $\underline{c} = \{1, 4\}$. In order to observe the sensitivity of IV/TSLS and other $\kappa-$class estimators to non–normality of stochastic component, we also generate disturbances from a multivariate t distribution with covariance matrix equal to $\boldsymbol{\Sigma}$, previously defined, using the following degrees of freedom: $DF = \{2, 6, 12\}$.

Non standard conditions and invalid instrument

In simulation study we consider a particular scenario whose validity condition is violated due to the presence of *weakly endogenous* instruments, which seems more realistic in practice. This means that we introduce a correlation between the instrument and structural disturbances. Recalling the just identified case, under a weakly endogenous instrument the system in (2.2, 2.3) can be expressed ad follows:

$$\begin{aligned} \mathbf{y} &= \mathbf{x}\beta + \boldsymbol{e} \\ \mathbf{x} &= \pi\mathbf{z} + \boldsymbol{v} \\ \boldsymbol{e} &= \gamma\mathbf{z} + \boldsymbol{u}, \end{aligned} \tag{2.33}$$

where $\boldsymbol{e}$ in (2.33) is constructed to ensure a non zero correlation between instrument and structural disturbances (of second stage). Moreover, the estimation error becomes:

$$\hat{\beta}_T^{IV} - \beta = (\mathbf{z}'\mathbf{x})^{-1}\mathbf{z}'\boldsymbol{e}.$$

The parameter $\gamma = \phi/\sqrt{T}$, where ϕ is a fixed constant, is modeled local to zero as the sample size increases. When $\phi = 0$, then $\gamma = 0$ and instruments are completely exogenous, as required in assumption of section 2.1.2. This type of modeling, introduced by Conley et al. (2012), which may be called *weakly endogenous* instrument asymptotics, will not be formally discussed because our main interest is focused on the performance of estimators and tests under a mutual failure in the basic assumpions. Different $\phi = \{0.1, 0.5, 1\}$ are chosen to observe the behaviour of IV estimator and its associated t/Wald statistic, depending on γ and π, under weakly endogenous (and weak) instrument. In fact, the asymptotic expected value of $\hat{\beta}^{IV}$ is equal to ϕ/π.

First stage and t–statistic

Another purpose is to quantify bias of IV/TSLS with respect to the degree of identification and level of endogeneity, comparing performances of first stage F tests and first stage R squared, denoted as R_f^2. We report population bias of IV/TSLS, expressed in (2.30), and its estimates, the estimation of (possibly average) concentration parameter, obtained as $\hat{\mu^2}/k = F - 1$, and finally the median of F statistics. We also compute mean and median of the first stage R_f^2 across $M = 100000$ replications. Proportion of F greater than 10, i.e. the empirical threshold proposed by Staiger and

Stock (1997), is reported, together with the frequency of F statistic greater than the following thresholds: $\{16.4, 10.83\}$, representing the critical values provided by Stock and Yogo (2005) when $k = 1, 5$, based on size distortion of t/Wald test (IV case) and on the relative bias of TSLS against OLS ($k = 5$). In particular, according to Stock et al. (2002), these values ensure that relative bias of TSLS with respect to OLS is less than the ten percent, under the assumption of normally distributed stochastic components. Furthermore, we consider DGP with multivariate t disturbances with six degrees of freedom.

Finally, we analyze performance of t–statistic for the null hypothesis of $H_0 : \beta = \beta_0$ associated to $\kappa-$class estimators (TSLS, LIML and two types of Fuller) under different degrees of endogeneity and when instruments are weakly endogenous, using DGP of 2.33, both in just–identified case and considering more than one instrument, setting everywhere $\beta_0 = 0$.

2.3.3 Main results of the simulation

Just identified case (IV)

In all of Tables we use bold to emphasize worst results and red colour to highlight the best performance according to selected indicators. Tables 2.1 and 2.2 show the main results of IV estimator: when $k = 1$ and the instrument is totally irrelevant, median of $\hat{\beta}_T^{IV}$ is practically equal to ρ; this appears because $plim(\hat{\beta}_T^{IV} - \beta) = plim(\hat{\beta}_T^{OLS}) = \rho$ in our setting. In terms of Median point estimates, IV performs well; its median across M generated samples is practically equal to zero for all $\mu^2 > 1$. MAE and RMSE decrease very fast with the strength of instrument. However, RMSE reaches very huge values even if $\mu^2 = 10$, especially in small samples ($T = 100$), due to the no–moment problem of IV estimator, combined with weak identification. Ouysse (2011) suggests to apply the "adjusted" RMSE in this situation, excluding the first and the last 5% of the observation. Coverage rates are often nearly to one under low endogeneity and weak instrument. Nonetheless, when $\rho > 0.5$, they could be very lower than 0.95 when instrument is poorly correlated with the endogenous regressor.

When the DGP allows to include weakly endogeneity of instruments, the asymptotic distribution is centered on the (wrong) value ϕ/π, and rapidely tends to infinity in the unidentified case ($\mu^2 = \pi = 0$). These results are presented in Tables 2.7 and 2.8. When $\phi = 0.1$, i.e. low endogeneity of instrument, IV estimator performs well in terms of Median, MAE, Coverage rates and KS, especially when $\rho \leq 0.75$. However, under high endogeneity of instruments ($\phi = 1$), confidence intervals are severely understated in terms of coverage rates even if the instrument is highly correlated with endogenous regressor. Thus, KS statistic reflects the fact that distributions are centered on the wrong value $\beta + \phi/\pi$. In this case, KS values are approximately greater than 0.3 for all considered scenarios.

Figure 2.1 shows the empirical density and Ecdf of standardized IV estimators under different degrees of identification, where red line represents the standard gaussian distribution. Furthermore, Figure 2.2 contains results about two identification scenarios. To summarize, high endogeneity combined with weak instruments and no moment problem could exacerbate non–normality of IV estimator.

Overidentified case

When the number of instruments increases, bias of TSLS may be severe even if the population level of identification is not deemed weak according to $F > 10$ rule of thumb. In particular, considering the case of $\rho = 0.9$ and $\mu^2/k \leq 20$, confidence intervals appear seriously understated. In fact, when $\mu^2/k = 20$, we observe that may be very far from the nominal level, as described in Tables 2.3 and 2.4, where it is also noticeable that median increases with respect to IV case, at the same level of μ^2/k, due to the overidentification.

Figure 2.3 shows empirical density of standardized TSLS estimator with $\rho = 0.9$, while in figure 2.4 we plot empirical densities of TSLS under weak and strong instruments scenarios for different degrees of endogeneity . When ρ takes high values, the empirical densities are far from standard normal even if $\mu^2/k = 10$, similarly to IV case. Regarding other $\kappa-$class estimators, Fuller with $\underline{c} = 1$ performs well in terms of coverage rates, reaching the $1 - \alpha$ level even if $\mu^2 = 10$ for each level of endogeneity. Moreover, it presents lower values for KS in case of $k = 5$ instruments, as shown in Table 2.6, while LIML (Table 2.5) also performs well, especially regarding Median point estimates, resulting less sensitive to the degree of endogeneity than TSLS and Fuller. This fact confirms that LIML results median unbiased when instruments are not irrelevant or very weak ($\mu^2/k = 1$). Finite sample behaviour of LIML and Fuller under different μ^2/k and ρ could be visualized in Figures 2.5, 2.6 2.7, and 2.8.

Hypothesis testing

When inference is conducted through conventional t/Wald test, associated to a $\kappa-$class estimator, the distribution of t-statistic under the null hypothesis could be non–normal in finite samples under weak or irrelevant instruments. Figure 2.9 presents asymptotic distribution of t–statistic (under the null hypothesis) when $\rho = 0.5$ (left panel) and $\rho = 0.9$ (right panel) for IV ($k = 1$) and TSLS estimator considering five instruments. These distributions are very different from standard normal even if $\mu^2 = 10$, under high endogeneity, and substantially skewed presenting heavy tail on the right. In Figure 2.11 we clearly notice that t/Wald statistics associated to LIML and Fuller estimators are closer to standard normal when $\mu^2 = 10$, exhibiting non–normality and skewness when instruments are very weak or irrelevant ($\mu^2/k \approx 0$). Nevertheless, LIML estimator seems to be more robust to the level of endogeneity (ρ), expecially when instruments are collectively considered not too weak ($\mu^2/k = 10$).

Table 2.12 shows the empirical size of t–Wald test associated to $\kappa-$class estimator under (jointly) normal disturbances and two different sample sizes, while OLS is considered as a worst case benchmark. The number of instruments is equal to 5 and sample size is equal to $T = 100, 1000$. Wald/t test associated to LIML estimator presents the best performance; TSLS case is again severely oversize in finite samples, especially under high endogeneity. Under jointly normal disturbances, Fuller with $\underline{c} = 1$ outperforms that with $\underline{c} = 4$, performing very poor under irrelevant and very weak instruments. Moreover, weakly endogeneity of instruments could dramatically affect performance empirical size, as viewed in Table 2.13. Under very low endogeneity of instruments, i.e. $\phi = 0.1$, LIML outperforms other estimators, while moderate endogeneity ($\phi = 1$) deteriorates rejection frequency for all considered methods. Furthermore, Figure 2.12 shows rejection frequency of t/Wald test for the just–identified case under different possible endogeneity levels; when instruments are weakly invalid ($\phi = 0.1$), rejection frequencies are close to the case of valid instruments for each identification level. When high invalidity arises, the test over–rejects too often, especially when $\phi = 1$.

Regarding first stage evidences and bias of TSLS under normal disturbances, presented in Table 2.14, we firstly notice that estimated bias (of IV/TSLS) is close to theoretical value even if instruments are not too strong, especially in the TSLS case. We also observe that both $F > 10$ rule of thumb and F test using critical values from Stock and Yogo perform better when $k = 5$. However, if $k = 1$ and $\rho = 0.9$, F test seems not to capture the presence of a large bias in the finite sample distribution of the estimator, occurring even if $\mu^2 = 10$. When disturbances are generated from a multivariate t with 6 degrees of freedom, estimated bias could be different from its approximated theoretical counterpart, as highlighted in Table 2.15, first stage R^2 is lower than that obtained under normal errors for every identification level, and also $\hat{\mu}^2 = E(F) - 1$ sistematically underestimates population μ^2. To summarize, F tests loses power even under moderately strong instruments under skewed and non–normal disturbances. In these cases is not trivial to recognize the true correlation between the instruments and endogenous regressor estimating the strength of instruments.

Finally, we compare rejection frequency First stage F rules under non–standard conditions applying different data generating processes. Figure 2.13 shows rejection frequency of both $F > 10$ test and $F > 16.4$ empirical power under weak instrument asymptotics with $k = 1$ and four different DGPs, including logNormal instrument and disturbances coming from a multivariate t with two and six degrees of freedom. To summarize, when disturbances are generated through a multivariate t with 2 degrees of freedom, the test presents very low power even under very strong instrument ($\mu^2 = 60$).

2.4 Bootstrap methods in instrumental variables

In this section, several bootstraps in linear instrumental variable are discussed and reviewed; different methods may be applied with respect to a specific inference of interest. A first possible application regards bias–adjustment of IV/TSLS estimators through bootstrap–based bias corretion. The basic idea is to estimate the bias of IV or $\kappa-$class estimators using the average (or possibly the median) of B bootstrap replications as a proxy for the expected value. The method is introduced in Hsu et al. (1986), and successively studied by Flores-Lagunes (2007) under the combination of weak instruments, high endogeneity and moderate degree of overidentification. Flores-Lagunes (2007) finds that bootstrap–based bias corrected version of some $\kappa-$class estimators, such as LIML and Fuller, may presents better performance than conventional IV/TSLS. Therefore, simulation study highlights critical issues in obtaining realiable point estimates under weak identification, especially combined with a moderate amount of instruments ($k = 30$) and high degrees of endogeneity. In the same context Ouysse (2013) applies two types of double bootstrap as a bias correction device, finding better performances by iterating the bootstrap or considering the method introduced by

Davidson and MacKinnon (2007). However, issues regarding bootstrap–based bias–correction are confirmed, especially when instruments are collectively weak, applying conventional TSLS. From a theoretical point of view, Chau (2014) proves the equivalence of indirect inference (II) and bootstrap–based bias correction of $\kappa-$class estimators.

Another topic of interest regards bootstrap methods in hypothesis testing. When inference is conducted through conventional t/Wald test, Moreira et al. (2009) prove invalidity of bootstrap under irrelevant instruments, meaning that the rejection probability under the null hypothesis does not tend to nominal level of the test as the sample size increases. Main issue is substantially related to non–pivotality of t/Wald statistic under weak or irrelevant instruments. Practically speaking it does not converge, in a bootstrap sense, to a free parameters distribution. To overcome these limitation, Davidson and MacKinnon (2008) propose two new types of residual–based bootstrap: the former is based on a consistent estimator of $\boldsymbol{\pi}$ introduced by Kleibergen (2002), while the latter essentially involves a bias correction on the estimated concentration parameter $\hat{\mu}^2$. These methods are also considered in the context of Sargan-type statistics (Davidson and Mackinnon, 2015), and bootstrapped confidence sets may be obtained inverting non–robust and robust tests (Davidson and MacKinnon, 2014).

Furthermore, bootstrap methods may be applied in the context of instruments selection; Wang et al. (2015) propose bootstrap approximation of mean square error (BMSE) using the analytical bias in order to choose the number of instrumental variables, while Inoue (2006) considers coverages errors estimated via bootstrap.

In the next paragraphs we discuss some parametric, semi–parametric and non–parametric bootstrap methods in IV linear models, summarizing some recent advances in hypothesis testing under weak instruments. Detailed algorithms of some methods are reported in the Appendix.

2.4.1 Parametric bootstrap

Parametric bootstrap method rely on the (possibly strictly) hypothesis of normally distributed and homoskedastic disturbances. In this case, the quantities $(u_t^\star, v_t^\star)$ are based on variance of the residuals and drawn from the following distribution:

$$\begin{pmatrix} u_t^\star \\ v_t^\star \end{pmatrix} \sim NID\left(0, \begin{bmatrix} \hat{\sigma}_u^2 & \hat{\sigma}_{uv} \\ \hat{\sigma}_{uv} & \hat{\sigma}_u^2 \end{bmatrix}\right).$$

Given $(u_t^\star, v_t^\star)$, the bootstrap data $D_T^\star = (y_t^\star, x_t^\star)$ are rebuilt in the following way:

$$\mathbf{y}^\star = \mathbf{x}^\star \hat{\beta}_T + \mathbf{u}^\star \tag{2.34}$$

$$\mathbf{x}^\star = \mathbf{Z}\hat{\pi}_t + \mathbf{v}^\star. \tag{2.35}$$

Considering quantities constructed in (2.34, 2.35), the parametric bootstrapped TSLS estimator becomes:

$$\begin{aligned}\hat{\beta}_T^{TSLS\star} &= (\mathbf{x}^{\star\prime} P_Z \mathbf{x}^\star)(\mathbf{x}^{\star\prime} P_Z \mathbf{y}^\star) \\ &= \hat{\beta}_T^{TSLS} + (\mathbf{x}^{\star\prime} P_Z \mathbf{x}^\star)(\mathbf{x}^{\star\prime} P_Z \mathbf{u}^\star),\end{aligned}$$

while IV bootstrap counterpart is:

$$\begin{aligned}\hat{\beta}_T^{IV\star} &= (\mathbf{z}'\mathbf{x}^\star)^{-1}\mathbf{z}'\mathbf{y}^\star = \frac{\sum_{t=1}^T z_t y_t^\star}{\sum_{t=1}^T z_t x_t^\star} \\ &= \hat{\beta}_T^{IV} + \frac{\sum_{t=1}^T z_t u_t^\star}{\sum_{t=1}^T z_t x_t^\star}.\end{aligned}$$

Furthermore, it may be possible to assume jointly normality of disturbances and instruments resampling the z_t. For example in IV case $z_t^\star \sim N(0, \hat{\sigma}_z^2)$, assuming $E^\star(z_t u_t) = E^\star(z_t v_t) = 0$.

Another parametric bootstrap is the fixed regressor i.i.d. method, in which all regressors (from both first and second stage) take their original values in the bootstrap DGP, assuming normality of structural disturbances. So, considering the just identfied case where $u_t^\star \sim N(0, \hat{\sigma}_v^2)$ and $y_t = \hat{\pi}_T x_t + u_t^\star$, the IV estimator takes the following form:

$$\hat{\beta}_T^{IV\star} = \frac{\sum_{t=1}^T z_t y_t^\star}{\sum_{t=1}^T z_t x_t} = \hat{\beta}_T^{IV} + \frac{\sum_{t=1}^T z_t u_t^\star}{\sum_{t=1}^T z_t x_t}.$$

However, this method is discouraged in IV framework because it does not involve quantities coming from first stage that are affected by the strength of the instruments. Furthermore, illustrated parametric methods are not recommended if the joint normality of the stochastic components is not guarantee, or safely rejected.

2.4.2 Non parametric method: Pair bootstrap

The first bootstrap method applied in IV setting is the pair (also called paris or pairwaise) bootstrap, introduced by Freedman (1984). This method essentially involves the sampling of endogenous and exogenous variables, i.e. the rows of matrix: $(\mathbf{y}, \mathbf{x}, \mathbf{Z}) = (y_t, x_t, Z_{k,t})_{t=1}^T$. Then, the bootstrap counterpart of TSLS estimator in (2.9) takes the following form:

$$\hat{\beta}_T^{TSLS*} = (\mathbf{x}^{*\prime} P_{Z^*} \mathbf{x}^*)^{-1} \mathbf{x}^{*\prime} P_{Z^*} \mathbf{y}^*, \tag{2.36}$$

where $P_{Z^*} = \mathbf{Z}^*(\mathbf{Z}^{*\prime}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*\prime}$. In the just identified case, given the resampled observation of the original data $(y_t^*, x_t^*, z_t^*)'$, the bootstrap estimators becomes:

$$\hat{\beta}_T^{IV*} = (\mathbf{z}^{*\prime}\mathbf{x})^{-1} \mathbf{z}^{*\prime}\mathbf{y} = \frac{\sum_{t=1}^T z_t^* y_t^*}{\sum_{t=1}^T z_t^* x_t^*}. \tag{2.37}$$

As pointed out by Freedman et al. (1984), Flores-Lagunes (2007) and Wang et al. (2015), this method could be slightly modified to avoid some issues. Basically, the pair bootstrap does not guarantee orthogonality between the TSLS residuals, $\hat{u}_t = y_t - \hat{\beta}_T^{TSLS} x_t$, and instruments Z_t in the bootstrap world, i.e. $E^*(u_t^*, Z_t)$ could be different from zero. For this purpose, the vector of modified residuals $\tilde{\mathbf{u}} = M_Z \hat{\mathbf{u}}$, where $M_Z = I - P_Z$, is implemented in the resampling scheme, by using the matrix $W_{k,i} = (\tilde{\boldsymbol{u}}, \mathbf{x}, \mathbf{Z}) = (\tilde{u}_t, x_t, Z_{k,t})_{t=1}^T$ to obtain $(\tilde{u}_t^*, x_t^*, Z_t^*)$. The outcome variable could be reconstructed in the bootstrap world as $\tilde{y}_t^* = \hat{\beta}_T x_t^* + \tilde{u}_t^*$. Finally, the new bootstrap TSLS estimator is:

$$\hat{\beta}_T^* = (\mathbf{x}^{*\prime} P_{Z^*} \mathbf{x}^*)^{-1} \mathbf{x}^{*\prime} P_{Z^*} \tilde{\mathbf{y}}^*, \tag{2.38}$$

while in the perfectly identified case the bootstrap counterpart of IV estimator is: $\hat{\beta}_T^{IV*} = (\mathbf{z}^{*\prime}\mathbf{x}^*)^{-1} \mathbf{z}^{*\prime}\tilde{\mathbf{y}}^*$.

2.4.3 Semi–parametric method: Residual bootstrap

A more efficient approach is represented by residual bootstrap. This method is semi–parametric and requires assumptions of homoskedasticity and incorrelation of the stochastic components, without specifying a probability law but resampling residuals from its empirical (cumulative) distribution function (EDF). The first method is called unrestricted residual bootstrap (Davidson and MacKinnon, 2010) and consists in resampling residuals coming from the estimated model, defined as follows:

$$\begin{aligned} \hat{\mathbf{u}} = &\quad \mathbf{y} - \hat{\beta}_T \mathbf{x} \\ \hat{\boldsymbol{v}} = &\quad \mathbf{x} - \mathbf{Z}\hat{\pi}_T, \end{aligned} \tag{2.39}$$

where $\hat{\beta}_T = \hat{\beta}_T^{IV/TSLS}$ is the TSLS or IV estimates and $\hat{\pi}_T = \hat{\pi}_T^{OLS}$ is the OLS estimates for $\boldsymbol{\pi}$ in the first (stage) equation[1]. The quantities in (2.39) are recentered to be $(\tilde{u}_t, \tilde{v}_t)$ with zero mean and *orthogonal* to the instruments. The disturbances[2] for bootstrap DGP are sampled from the following joint empirical distribution function:

$$\begin{pmatrix} u_t^* \\ v_t^* \end{pmatrix} \sim EDF \begin{bmatrix} \tilde{u}_t \\ \tilde{v}_t \end{bmatrix}.$$

Given new resampled disturbances, bootstrap DGP may be written in the following way:

$$\begin{aligned} \mathbf{y}^* &= \mathbf{x}^* \hat{\beta}_T + \mathbf{u}^* \\ \mathbf{x}^* &= \mathbf{Z}\hat{\pi}_T + \boldsymbol{v}^*. \end{aligned} \tag{2.40}$$

Finally, the bootstrap counterpart of the TSLS estimator is computed as follows:

$$\begin{aligned} \hat{\beta}_T^{TSLS*} &= \left(\mathbf{x}^{*\prime} P_Z \mathbf{x}^*\right)^{-1} \left(\mathbf{x}^{*\prime} P_Z \mathbf{y}^*\right) \\ &= \left[\mathbf{x}^{*\prime}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{x}^*\right]^{-1} \left[\mathbf{x}^{*\prime}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}^*\right]. \end{aligned} \tag{2.41}$$

In the perfectly identified case, given $\hat{y}_t^* = x_t^* \hat{\pi}_T + u_t^*$ and $x_t^* = z_t \hat{\pi}_T + v_t^*$, where $(u_t^*, v_t^*)' \sim EDF(\hat{v}_t, \hat{u}_t)'$, the bootstrap estimator in (2.41) is reduced to:

$$\hat{\beta}_T^{IV*} = (\mathbf{z}'\mathbf{x}^*)^{-1}\mathbf{z}'\mathbf{y}^* = \frac{\sum_{t=1}^T z_t y_t^*}{\sum_{t=1}^T z_t x_t^*}. \tag{2.42}$$

[1]We point out the difference between resampled residuals, denoted as u^* and v^*, from quantities of Section 2.4.1 indicated with $u^\star$ and $v^\star$, coming from a specific probability law.

[2]Davidson and MacKinnon (2008) suggest to multiply residuals v_t by a factor equal to $(T/(T-k))^{1/2}$.

Resampled instruments

In residual–based bootstrap, instruments $\mathbf{Z}$ can be assumed as fixed, as previously shown, or stochastic. The latter assumption requires the resampling of instrumental variables in the bootstrap DGP, as proposed for example in work of Moreira et al. (2009). In this case, the vector $\mathbf{Z}$ may be substituted with $\mathbf{Z}^*$.; then the second equation of (2.39) becomes:

$$\mathbf{x}^* = \mathbf{Z}^*\hat{\pi}_T + \mathbf{v}^*, \tag{2.43}$$

and bootstrapped TSLS estimator in (2.41), using (2.43), modifies to:

$$\hat{\beta}_T^{TSLS*} = \left(\mathbf{x}^{*\prime} P_{Z^*}\mathbf{x}^*\right)^{-1}\left(\mathbf{x}^{*\prime} P_{Z^*}\mathbf{y}^*\right), \tag{2.44}$$

where $P_{\mathbf{Z}^*} = \mathbf{Z}^*(\mathbf{Z}^{*\prime}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*\prime}$. Moreover, in the just identified case the expression in (2.43) is simplified to $x_t^* = \hat{\pi}_T z_t^* + v_T^*$, and bootstrap counterpart IV estimator, with resampling instrument, may be written as:

$$\hat{\beta}_T^{IV*} = (\mathbf{z}^{*\prime}\mathbf{x}^*)^{-1}\mathbf{z}^{*\prime}\mathbf{y}^* = \frac{\sum_{t=1}^T z_t^* y_t^*}{\sum_{t=1}^T z_t^* x_t^*}. \tag{2.45}$$

Davidson and MacKinnon (2010) include this method in the pairwise–type bootstraps, because it involves resampling of instruments and projections of endogenous variables, e.g. (u_t^*, v_t^*, Z_t^*). Nevertheless, we denote it as residual bootstrap with resampled instruments.

Therefore, it may be useful to remark that IV and OLS residuals have 0 mean if a constant term is included in the model, in both equations. In this case the recentering of residuals is not necessary and basically $(\hat{u}_t, \hat{v}_t)' = (\hat{u}_t, \hat{v}_t)'$

Hypothesis Testing

The previous methods may be straightforward adapted to hypothesis testing, without imposing the null in the bootstrap DGP. A first application regards bootstrap counterpart of t/Wald statistic for testing the null hypothesis $H_0 : \beta = \beta_0$; the test statistics takes the following form:

$$\tau_T^* = \frac{\hat{\beta}_t^* - \hat{\beta}_T}{\tilde{\omega}_T^*/\sqrt{T}}, \tag{2.46}$$

where $\hat{\omega}^*/\sqrt{T}$ is the bootstrap counterpart of $\hat{\omega}/\sqrt{T}$, i.e. the standard error of $\hat{\beta}_T$. The associated bootstrap p–value is based on the assumptions that limiting distribution of τ_T^*, i.e. τ_∞ is symmetric, and could be computed as the proportion of bootstrap statistics greater than the estimated one:

$$p^* = B^{-1}\sum_{b=1}^{B}(\left|\tau_{T,b}^*\right| \geq |\tau_T|). \tag{2.47}$$

Given the expression in (2.47), the null hypothesis is safely rejected if $p^* \leq \alpha$, where α is the I type error level, for example $\alpha = 0.05$ or $\alpha = 0.01$.

However, different methods are developed imposing the null hypothesis in the bootstrap DGP. In order to test null of $H_0 : \beta = 0$ in TSLS estimation, the bootstrap sample could be constructed, imposing $\beta = \beta_0$, as follows:

$$\begin{aligned} \hat{\mathbf{u}}(\beta_0) = & \ \mathbf{y} - \beta_0 \cdot \mathbf{x} \\ \hat{\boldsymbol{v}} = & \ \mathbf{x} - \mathbf{Z}\hat{\pi}_T, \end{aligned} \tag{2.48}$$

where $\hat{\mathbf{u}}(\beta_0)$ is the vector of residuals from the structural equation, induced by imposing the null hypothesis. Hence, Bootstrap DGP is constructed using sampled residuals from:

$$\begin{pmatrix} u_t^* \\ v_t^* \end{pmatrix} \sim EDF \begin{bmatrix} \tilde{u}(\beta_0)_t \\ \tilde{v}_t \end{bmatrix},$$

and bootstrapped data:

$$x_t^* = Z_t\hat{\pi}_T + v_t^* \tag{2.49}$$

$$y_t^* = \beta_0 x_t + u_t^*. \tag{2.50}$$

Finally, the bootstrap counterpart of t–statistic is computed in the following way:

$$\tau_T^* = \sqrt{T}\frac{(\hat{\beta}_T^*(\beta_0) - \beta_0)}{\hat{\omega}^*(\beta_0)} \tag{2.51}$$

where $\hat{\beta}_T^*(\beta_0)$ and its standard error $\hat{\omega}^*(\beta_0)T^{-1/2}$ are estimated through IV/TSLS [3] using quantities in (2.49) and (2.50). This method is also called *restricted* residual bootstrap.

[3] Or another $\kappa-$class estimator

Heteroskedasticity

Among the previoulsy introduced methods, only the pair bootstrap is considered valid even if the disturbances are not (jointly) homoskedastic. Nevertheless, residual–based techniques could be modified to take account of an heteroskedasticity of unknown form using the so called "wild" residuals, introduced by Wu (1986). The method, named Wild bootstrap, is based on a transformation of the disturbances which consists in multiplying them by a random variable with zero mean and variance equal to unit. Considering the residuals in (2.39), the method involves resampling from:

$$\begin{pmatrix} u_t^{*w} \\ v_t^{*w} \end{pmatrix} \sim EDF \begin{bmatrix} \tilde{u}_t \xi_t \\ \tilde{v}_t \xi_t \end{bmatrix},$$

where the auxiliary random variable ξ_t has the following properties: $E(\xi_t) = 0$ and $V(\xi_t) = 1$. Common choices for are the Mammen distribution (Mammen, 1993) and the Rademacher distribution: $\xi_t \in \{-1, 1\}$ where $P(\xi_t = -1) = P(\xi_t = 1) = 1/2$. Finally, bootstrap data are reconstructed using the new $(u_t^{*w} v_t^{*w})'$ instead of $(u_t^*, v_t^*)'$, for example using the same syste viewed in the system (2.40). This method may be straightforward applied in hypothesis testing.

2.4.4 Recent improvements in bootstrap methods

Davidson and MacKinnon (2008) introduce new residual–based bootstrap methods producing better performance, with respect to conventional (parametric and non–parametric) methods, under weak instruments and certain degree of overidentification (they consider 11 instrumental variables). The basic idea regards issue related to inefficiency of OLS estimator for $\boldsymbol{\pi}$ when instruments are collectively weak; to overcome this problem they apply a more efficient estimator for $\tilde{\pi}_T$, introduced by Kleibergen (2002). In the so–called Unrestricted (residual) Efficient bootstrap the equation in (2.3) becomes:

$$\mathbf{x} = \mathbf{Z}\boldsymbol{\pi} + \delta \hat{\mathbf{u}} + \boldsymbol{\varepsilon}, \tag{2.52}$$

where $\boldsymbol{\varepsilon}$ is the $T \times 1$ vector of the errors, δ is a scalar parameter and $\hat{\mathbf{u}}$ is the vector containing residuals, computed using $\hat{\beta}_T^{TSLS}$ in IV regression. Following expression (2.52), OLS estimates are $\tilde{\boldsymbol{\pi}}$ and $\tilde{\delta}$, and then the residuals implemented in the bootstrap procedures are:

$$\tilde{\boldsymbol{v}}^{ur} = \mathbf{x} - \mathbf{Z}\tilde{\boldsymbol{\pi}}, \tag{2.53}$$

and corresponds to the OLS residual augmented by the quantity $\hat{\delta}\hat{\mathbf{u}}$. Using procedure described in (2.52) and residuals in (2.53), estimator of the parameters $\beta, \boldsymbol{\pi}$ is asymptotically equivalent to Three Stage Least Squares (3SLS) if β is estimated through TSLS[4]. With the procedure described in (2.52), pairs of resampled residuals are the following:

$$\begin{pmatrix} u_t^* \\ v_t^* \end{pmatrix} \sim EDF \begin{pmatrix} \tilde{u}_t \\ \tilde{v}_t^{ur} \end{pmatrix}. \tag{2.54}$$

If the null hypothesis is imposed, i.e. $\beta = \beta_0$, the system becomes:

$$\begin{aligned} \mathbf{y} &= \mathbf{x}\beta_0 + \mathbf{u} \\ \mathbf{x} &= \mathbf{Z}\boldsymbol{\pi} + \delta\hat{\mathbf{u}}(\beta_0) + \boldsymbol{\varepsilon}, \end{aligned} \tag{2.55}$$

This method is called restricted (residual) efficient bootstrap (RE), and the expression for the τ_T^* is equal to the (2.51). If $\beta_0 = 0$, then $\hat{\mathbf{u}}(\beta_0)$ reduced to $\delta M_Z\mathbf{y}$, and $\tilde{\mathbf{v}}$ is equal to OLS residuals.

Finally summarize some recent advances in this topic. In the context of clustered–robust inference, Finlay and Magnusson (2014) propose a wild cluster constrained residual–based bootstrap, which outperforms other resampling methods in hypothesis testing under clustered disturbances, even if instruments are weak. Moreover, Wang and Kaffo (2016) demostrate the invalidity of residual efficient bootstrap under many instruments asymptotics for TSLS and LIML estimator. They modifies this algorithm in order to obtain validity bootstrapping LIML and Fuller estimators under the many instruments sequence. Finally, Doko Tchatoka (2015) demonstrates the validity of bootstrap in the context of DWH (Durbin–Wu–Hausman) endogeneity tests even under weak or irrelevant instruments.

2.5 Concluding Remarks

In this chapter we review instrumental variable estimation under weak instrument asymptotics. IV/TSLS and other estimators, nested in the so called $\kappa-$class, are

[4] If LIML is used instead of TSLS, estimator is asymptotically equivalent to the so called Full Information Maximum Likelihood (FIML)

typically used to obtain structural coefficient associated to an endogenous regressor, considering a fixed number of instrumental variables equal to k. Conventional asymptotic approximation works if validity of instruments, including exogeneity and relevance conditions, holds. Under weak instrument asymptotics, estimators are no longer consistent and, in particular, normality of IV/TSLS (but also of other $\kappa-$class estimators) is reflected in their finite samples properties. Furthermore, inference conducted through conventional t/Wald test, under null hypothesis $H_0 : \beta = \beta_0$, leads to misleading results. In particular, rejections frequencies of Wald test are too high when poorly relevant instruments are combined with moderate and high endogeneity. Moreover, first stage F test for detecting weak instruments could presents very low power in non normal homoskedastic cases, especially when the disturbances are joitly drawn from a multivariate non–gaussian distribution. The effect of weak identification can also be combined with high endogeneity and overidentification, affecting finite sample bias.

Since Monte Carlo simulation of $\kappa-$class estimators are computationally straigthforward, we argue the usage of KS distance to evaluate performance in the context of IV and GMM estimators under weak/invalid instruments and non–standard conditions. This quantity could be helpful to quantify severity of non–normality under different DGPs. Our results confirm that LIML presents good performance in terms of median point and coverage rates, especially under homoskedastic and normally distributed disturbances. However, when instrumens are too weak it presents huge variance and unreliable standard errors.

We also discuss bootstrap methods in IV estimation: they could be parametric, non–parametric and semi–parametric. Former methods require strictly assumptions on the distribution of disturbances (e.g. joint normality), while semi–parametric and non–parametric bootstrap are based on the resampling from empirical distribution of the stochastic components or variables (pair bootstra casep). Semi–parametric methods refer to bootstrap DGP including estimates from both first and second stage. Some of these methods are valid even in presence of heteroskedasticity (Wild and Pair) of unknown form, while others (Residual Efficient) are able to improve performance of tests imposing the null in bootstrap DGPs.

2.6 Appendix

Weak instrument asymptotics

Just identified case

In the just–identified case, the first stage parameter $\pi = \pi(T) = c \cdot T^{-1/2}$ while the variance of instrument is equal to $Q_{zz} = \sigma_z^2$. The distribution of $\hat{\pi}_T^{OLS}$ under weak instrument asymptotics is the following:

$$\begin{aligned}
\hat{\pi}_T^{OLS} &= (\mathbf{z}'\mathbf{z})^{-1}\mathbf{z}'\mathbf{x} = \pi + (\mathbf{z}'\mathbf{z})^{-1}\mathbf{z}'\mathbf{v} \\
&= \pi + \frac{\sum_{t=1}^T z_t v_t}{\sum_{t=1}^T z_t^2} = \frac{c}{\sqrt{T}} + \frac{\sum_{t=1}^T z_t v_t}{\sum_{t=1}^T z_t^2}; \\
\sqrt{T}\hat{\pi}_T &= c + \sqrt{T}\frac{\sum_{t=1}^T z_t v_t}{\sum_{t=1}^T z_t^2} = c + \frac{\sum_{t=1}^T z_t v_t/\sqrt{T}}{\sum_{t=1}^T z_t^2/T} \\
&\xrightarrow{d} N\left(c, \frac{\sigma_v^2}{\sigma_z^2}\right).
\end{aligned}$$

Moreover, the asymptotic distribution of $\hat{\beta}_T^{IV}$ under a single weak instrument becomes:

$$\begin{aligned}
\hat{\beta}_T^{IV} - \beta &= \frac{T^{-1}\sum_{t=1}^T z_t u_t}{T^{-1}\sum_{t=1}^T z_t x_t} = \frac{T^{-1}\sum_{t=1}^T z_t u_t}{T^{-1}\sum_{t=1}^T z_t(c/\sqrt{T}z_t + v_t)} \\
&= \frac{T^{-1}\sum_{t=1}^T z_t u_t}{T^{-1/2}c\sum_{t=1}^T z_t^2 + T^{-1}\sum_{t=1}^T z_t v_t} \\
&\xrightarrow{d} \frac{w_{zu}}{(c\sigma_z^2 + w_{zv})}.
\end{aligned}$$

We recall that under this nesting $\hat{\beta}_T^{IV}$ is a non–normally distributed if $c < \infty$. Hence, this distribution may be rewritten as follows:

$$\begin{aligned}
\frac{w_{zu}}{(c\sigma_z^2 + w_{zv})} &= \frac{N(0, \sigma_z^2\sigma_u^2)}{c\sigma_z^2 + N(0, \sigma_z^2\sigma_v^2)} \\
&= \frac{\sigma_z^2\sigma_u^2 N(0,1)}{c\sigma_z^2 + \sigma_z^2\sigma_u^2 N(0,1)} = \frac{N(0,\sigma_u^2)}{c + N(0,\sigma_v^2)}.
\end{aligned}$$

Assuming $\sigma_z = 1$ and $c^2 = \mu^2$, previous expression can be simplified to:

$$\hat{\beta}_T^{IV} - \beta \xrightarrow{d} \frac{w_{zu}}{(\mu + w_{zv})} = \frac{N(0,\sigma_u^2)}{\mu + N(0,\sigma_v^2)}.$$

In special unidentified case where $c = \mu = \pi = 0$, IV estimator converges to:

$$\hat{\beta}_T^{IV} - \beta \xrightarrow{d} \frac{w_{zu}}{w_{zv}} = \frac{N(0,\sigma_u^2)}{N(0,\sigma_v^2)}.$$

Overidentified Case

When $k > 1$ the vector of first stage (nuisance) paremeters $\boldsymbol{\pi_0} = C/\sqrt{(T)}$. Again, OLS estimator converges as $T \to \infty$ to the following expression:

$$\begin{aligned}
\hat{\pi}_T^{OLS} &= (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{x} = \boldsymbol{\pi_0} + (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{v} \\
&= T^{-1/2}C + (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{v} \\
\sqrt{T}(\hat{\pi}_T) &= \sqrt{T}\left(T^{-1/2}C + (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{v}\right) \\
&= C + \sqrt{T}\left[(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{v}\right] \\
&= C + \left[\frac{(\mathbf{Z}'\mathbf{Z})^{-1}}{T}\frac{\mathbf{Z}'\mathbf{v}}{\sqrt{T}}\right] \\
&\xrightarrow{d} N_k(C, \sigma_v^2\mathbf{Q}_{ZZ}^{-1}).
\end{aligned}$$

Furthermore, the TSLS estimator $\hat{\beta}_T^{TSLS}$ is:

$$\begin{aligned}
\hat{\beta}_T^{TSLS} &= (\mathbf{x}'P_Z\mathbf{x})^{-1}(\mathbf{x}'P_Z\mathbf{y}) \\
&= \left[\mathbf{x}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{x}\right]^{-1}\left[\mathbf{x}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}\right] \\
&= \beta + (\mathbf{x}'P_Z\mathbf{x})^{-1}(\mathbf{x}'P_Z\mathbf{u}),
\end{aligned} \tag{2.56}$$

where denominator of expression (2.56) may be rewritten as:

$$\begin{aligned}
\mathbf{x}'P_Z\mathbf{x} &= (\mathbf{x}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{x}) \\
&= (\mathbf{Z}\boldsymbol{\pi_0} + \mathbf{v})'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{Z}\boldsymbol{\pi_0} + \mathbf{v}) \\
&= \left(\frac{(\mathbf{Z}\boldsymbol{\pi_0} + \mathbf{v})'\mathbf{Z}}{\sqrt{T}}\right)\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{-1}\left(\frac{\mathbf{Z}'(\mathbf{Z}\boldsymbol{\pi_0} + \mathbf{v})}{\sqrt{T}}\right) \\
&\xrightarrow{d} (\mathbf{Q}_{ZZ}C + W_{Zv})'\mathbf{Q}_{ZZ}^{-1}(\mathbf{Q}_{ZZ}C + W_{Zv}).
\end{aligned}$$

Finally, numerator of (2.56) is:

$$\begin{aligned}
\mathbf{x}'P_Z\mathbf{u} &= (\mathbf{x}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{u}) = \frac{(\mathbf{x}'\mathbf{Z})}{\sqrt{T}}\frac{(\mathbf{Z}'\mathbf{Z})^{-1}}{T}\frac{(\mathbf{Z}'\mathbf{u})}{\sqrt{T}} \\
&= \frac{(\mathbf{Z}\boldsymbol{\pi_0}+\mathbf{v})'\mathbf{Z}}{\sqrt{T}}\frac{(\mathbf{Z}'\mathbf{Z})^{-1}}{T}\frac{\mathbf{Z}'\mathbf{u}}{\sqrt{T}} \\
&= \frac{(\mathbf{Z}\boldsymbol{\pi_0}+\mathbf{v})'\mathbf{Z}}{\sqrt{T}}\frac{(\mathbf{Z}'\mathbf{Z})^{-1}}{T}\frac{\mathbf{Z}'\mathbf{u}}{\sqrt{T}} \\
&= \left(\frac{\mathbf{Z}'\mathbf{Z}}{T}C + \frac{\mathbf{v}'\mathbf{Z}}{\sqrt{T}}\right)\frac{(\mathbf{Z}'\mathbf{Z})^{-1}}{T}\frac{\mathbf{Z}'\mathbf{u}}{\sqrt{T}} \\
&\xrightarrow{d} (\mathbf{Q}_{ZZ}C + W_{Zv})'\mathbf{Q}_{ZZ}^{-1}W_{Zu}.
\end{aligned}$$

Under special case of irrelevant instruments, i.e. $\boldsymbol{\pi_0} = \mathbf{0} = \mu^2 = 0$, the distribution of TSLS becomes:

$$\beta^{TSLS} - \beta \xrightarrow{d} \frac{W'_{Zv}W_{Zu}}{W'_{Zv}W_{Zu}} = \frac{W_{Zu}}{W_{Zv}}.$$

Concentration parameter under strong and weak instruments

Under strong instrument asymptotics $\boldsymbol{\pi} = \boldsymbol{\pi_0}$, the concentration parameter $\mu^2 \xrightarrow{p} \infty$ as $T \to \infty$. For the IV case with $\pi = \pi_0$ the following asymptotic result holds:

$$\mu^2 = T \cdot \frac{\sigma_z^2\pi_0^2}{\sigma_v^2} \xrightarrow{p} \infty.$$

When $k > 1$ under strong instrument asymptotics, $\boldsymbol{\pi} = \boldsymbol{\pi}_0$, we have:

$$\mu^2 = T \cdot \frac{\boldsymbol{\pi_0}'\mathbf{Q}_{ZZ}\boldsymbol{\pi_0}}{\sigma_v^2} \xrightarrow{p} \infty.$$

To summarize, $\mu^2 = O_p(T)$ in both cases. Nevertheless, under weak instrument asymptotics, i.e. $\boldsymbol{\pi_0} = CT^{-1/2}$, the concentration parameter converges to a constant as $T \to \infty$:

$$\mu^2 = \frac{\boldsymbol{\pi_0}'\mathbf{Z}'\mathbf{Z}\boldsymbol{\pi_0}}{Tk\sigma_v^2} \xrightarrow{p} \frac{C'\mathbf{Q}_{ZZ}C}{k\sigma_v^2}.$$

This expression is simplified in the just–identified case:

$$\mu^2 = \frac{\pi^2\sigma_z^2}{\sigma_v^2} \xrightarrow{p} \frac{c^2\sigma_z}{\sigma_v^2}.$$

An R code to simulate TSLS and other $\kappa-$class estimators

Here we present the R code (R Core Team, 2016) to simulate the asymptotic distribution of IV/TSLS, $\kappa-$class estimators and their associated t–statistics (under the null hypothesis) under weak instruments asymptotics. The DGP presents jointly normally distributed and valid instruments, as indicated in Section 2.3.2. The output of the function written above includes the simulated standardized distribution of four $\kappa-$class estimators, i.e. TSLS, LIML, Fuller(1) and Fuller(4), and related t–statistics. We use the following R packages: *MASS* for the multi–normal distribution, *AER* for the estimation of IV/TSLS and *ivmodel* for other $\kappa-$class estimators.

```
library(MASS); library(AER); library(ivmodel)
kclass<-function(M,t,mu2,rho,k,beta){
# Inizialize estimators and standard errors
tsls<-numeric(); fuller<-numeric(); liml<-numeric()
fuller4<-numeric(); sd_full4<-numeric()
sdt<-numeric(); sd_liml<-numeric(); sd_full<-numeric()
I <- diag(1,k)
## Covariance matrix for the disturbances
Sigma<-matrix(c(1,rho,rho,1),2,2)
i_k=matrix(1,nrow=k,ncol=1);
## Weak instrument asymptotics and unidentification
if(mu2==0) pi<-i_k*0 else pi<-i_k%*%sqrt(mu2)/sqrt(t)
if(mu2==0) sqad<-NA else sqad<-sqrt(solve(t(pi)%*%pi)/(t))
for(i in 1:M){
 uv <- mvrnorm(t,c(0,0),Sigma)
 u<-uv[,1]; v<-uv[,2];
## DGP:
 z <- mvrnorm(t, rep(0,k) ,I)
 x <-z%*%pi+v;
 y <-beta*x+u
## Estimation:
 iv<-ivreg(y~x|z)
 tsls[i] <-iv$coef[2]
 liml[i]<-suppressWarnings(ivmodel(y,x,z)$LIML$point.est)
```

```
  fuller[i]<-suppressWarnings(ivmodel(y,x,z)$Fuller$point.est)
  fuller4[i]<-suppressWarnings(Fuller(ivmodel(y,x,z),b=4)$point.est)
  ols<-lm(x~z)$coefficients[2:(k+1)]
  sdt[i]<-summary(iv)$coefficients[2,2]
  sd_liml[i]<-suppressWarnings(ivmodel(y,x,z)$LIML$std.err)
  sd_full[i]<-suppressWarnings(ivmodel(y,x,z)$Fuller$std.err)
  sd_full4[i]<-suppressWarnings(Fuller(ivmodel(y,x,z),b=4)$std.err)
}
tsls_st<-(tsls-beta)/as.numeric(sqad);
liml_st<-(liml-beta)/as.numeric(sqad)
fuller_st<-(fuller-beta)/as.numeric(sqad);
fuller_st4<-(fuller4-beta)/as.numeric(sqad)
t.test_tsls<-tsls/sdt
t.test_liml<-liml/sd_liml
t.test_full<-fuller/sd_full
t.test_full4<-fuller4/sd_full4
out<-list(tsls_st,liml_st,fuller_st,fuller_st4,
t.test_tsls,t.test_liml,t.test_full,t.test_full4)
names(out)<-c('tsls','liml','fuller','fuller4',
't_tsls','t_liml','t_fuller','t_fuller4')
return(out)
}
## Example 1) Weak instruments:
  w_instr<- kclass(M,t,mu2=1,rho,k,beta=0)
## Example 2) Strong Instruments:
  s_instr<-kclass(M,t,mu2=20,rho,k,beta=0)
```

Bootstrap Algorithms in just identified case

In this subsection we briefly illustrate some bootstrap algorithms exposed in section 2.4 considering the just–identified case of IV estimator: parametric bootstrap, pair bootstrap and unrestricted residual–based bootstrap both with fixed and resample z_t.

- **Parametric Bootstrap**

Step 1. Estimate parameters $(\hat{\beta}_T^{IV}, \hat{\pi}_T^{OLS})$ and obtain residuals $(\hat{u}_t, \hat{v}_t)$

Step 2. Sample $(u_t^\star, v_t^\star)' \sim N_2(0, \hat{\Sigma})$ where $\hat{\Sigma} = \begin{pmatrix} \hat{\sigma}_u^2 & \hat{\sigma}_{uv} \\ \hat{\sigma}_{uv} & \hat{\sigma}_v^2 \end{pmatrix}$

Step 3. Reconstruct the bootstrap series as: $x_t^\star = \hat{\pi}_T^{OLS} z_t + v_t^\star; \quad y_t^\star = \hat{\beta}_T^{IV} x_t^\star + u_t^\star.$

Step 4. Bootstrap IV estimator: $\hat{\beta}_T^{IV\star} = (\mathbf{z}'\mathbf{x}^\star)^{-1}\mathbf{z}'\mathbf{y}^\star = \frac{\sum_{t=1}^T y_t^\star z_t}{\sum_{t=1}^T x_t^\star z_t}$

Step 5. Repeat 2–4 B times to obtain the distribution of bootstrapped IV estimator: $(\hat{\beta}_{T1}^{IV\star}, \ldots \hat{\beta}_{TB}^{IV\star})$.

- **Pair Bootstrap**

Step 1. Sample the endogenous variables and instrument: $(y_t^*, x_t^*, z_t^*)' \sim \text{EDF}(y_t, x_t, z_t)'$

Step 2. Compute bootstrap estimator $\hat{\beta}_T^{IV*} = (\mathbf{z}^{*\prime}\mathbf{x}^*)^{-1}\mathbf{z}^{*\prime}\mathbf{y}^* = \frac{\sum_{t=1}^T y_t^* z_t^*}{\sum_{t=1}^T x_t^* z_t^*}$.

Step 3. Repeat 1–2 B times to obtain the distribution of bootstrapped IV estimator: $(\hat{\beta}_{T1}^{IV*}, \ldots \hat{\beta}_{TB}^{IV*})$.

- **Residual Bootstrap** (resampled instrument)

Step 1. Estimate parameters $(\hat{\beta}_T^{IV}, \hat{\pi}_T^{OLS})$ and obtain $(\hat{u}_t, \hat{v}_t)$

Step 2. Sample residuals and instrument: $(u_t^*, v_t^*, z_t^*)' \sim \text{EDF}(\tilde{u}_t, \tilde{v}_t, z_t)'$

Step 3. Bootstrap DGP:

  - $x_t^* = \hat{\pi}_T^{OLS} z_t^* + v_t^*$;
  - $y_t^* = \hat{\beta}_T^{IV} x_t^* + u_t^*$.

Step 4. Bootstrap IV estimator: $\hat{\beta}_T^{IV*} = (\mathbf{z}^{*\prime}\mathbf{x}^*)^{-1}\mathbf{z}^{*\prime}\mathbf{y}^* = \frac{\sum_{t=1}^T y_t^* z_t^*}{\sum_{t=1}^T x_t^* z_t^*}$.

Step 5. Repeat steps from 2 to 4 B times in order to obtain $(\hat{\beta}_{T1}^{IV*}, \ldots \hat{\beta}_{TB}^{IV*})$.

- **Residual bootstrap** (fixed instrument):

Step 1. Estimate parameters $(\hat{\beta}_T^{IV}, \hat{\pi}_T^{OLS})$ and obtain residuals $(\hat{u}_t, \hat{v}_t)$

Step 2. Sample the residuals: $(u_t^*, v_t^*)' \sim \text{EDF}(\tilde{u}_t, \tilde{v}_t)'$

Step 3. DGP: $x_t^* = \hat{\pi}_T^{OLS} z_t + v_t^*; \quad y_t^* = \hat{\beta}_T^{IV} x_t^* + u_t^*.$

Step 4. $\hat{\beta}_T^{IV*} = (\mathbf{z}'\mathbf{x}^*)^{-1}\mathbf{x}^{*\prime}\mathbf{y}^*$.

Step 5. Repeat steps 2–4 B times to obtain $(\hat{\beta}_{T1}^{IV*}, \ldots \hat{\beta}_{TB}^{IV*})$.

Remark: pair bootstrap could be straightforward applied on all $\kappa-$class estimators because it does not involve the estimation of the parameters π, β in bootstrap DGP.

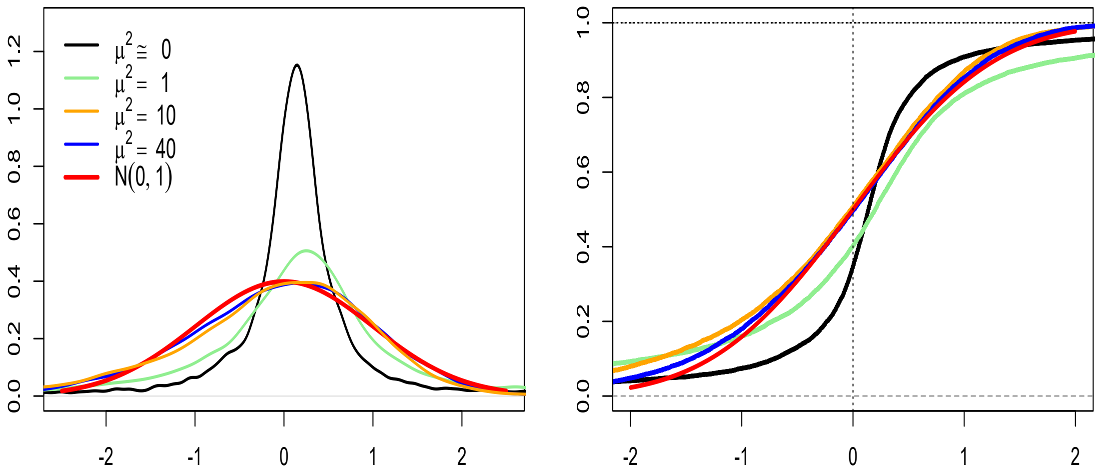
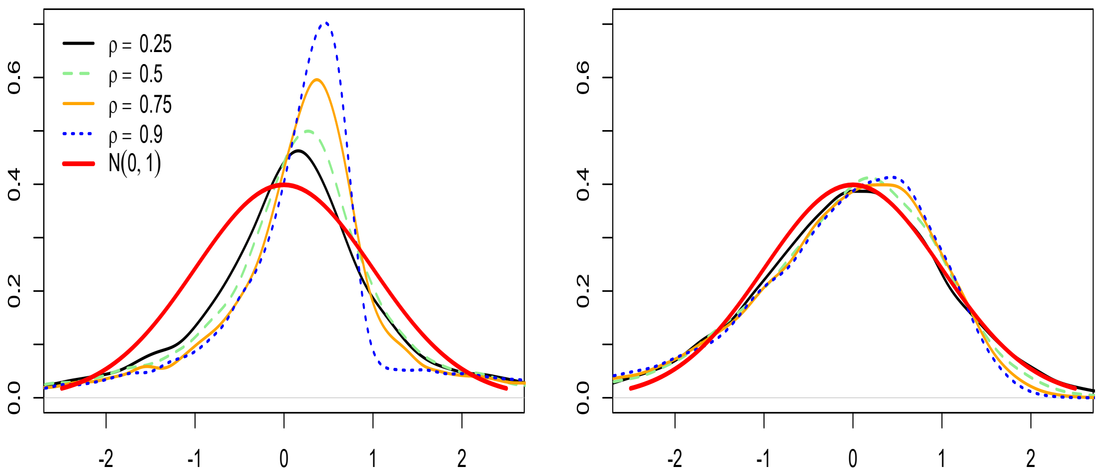


Figure 2.1: Empirical density and ECdf of standardized IV estimator under different degrees of identification.



(a) $\mu^2 = 1$ (b) $\mu^2 = 20$

Figure 2.2: Empirical density of standardized IV estimator under different levels of endogeneity for $\mu^2 = 1, 20$.

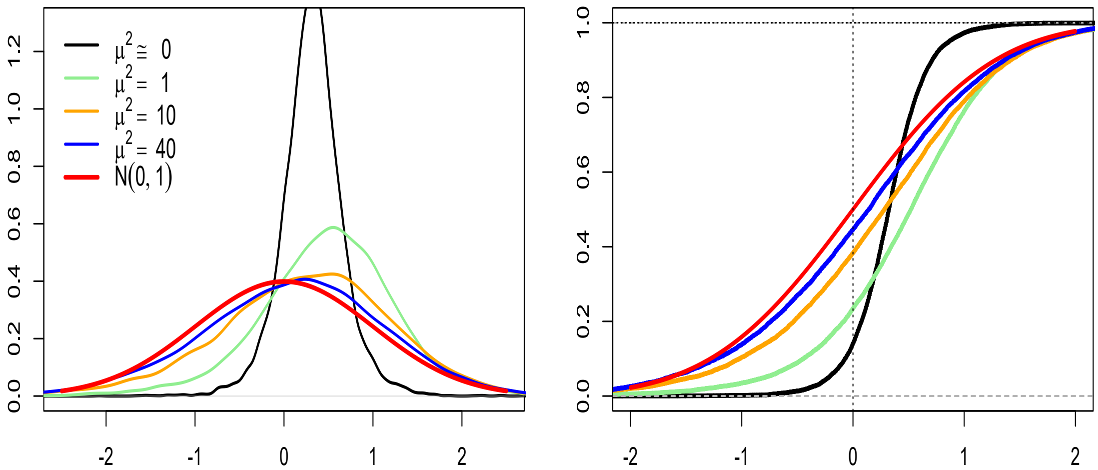
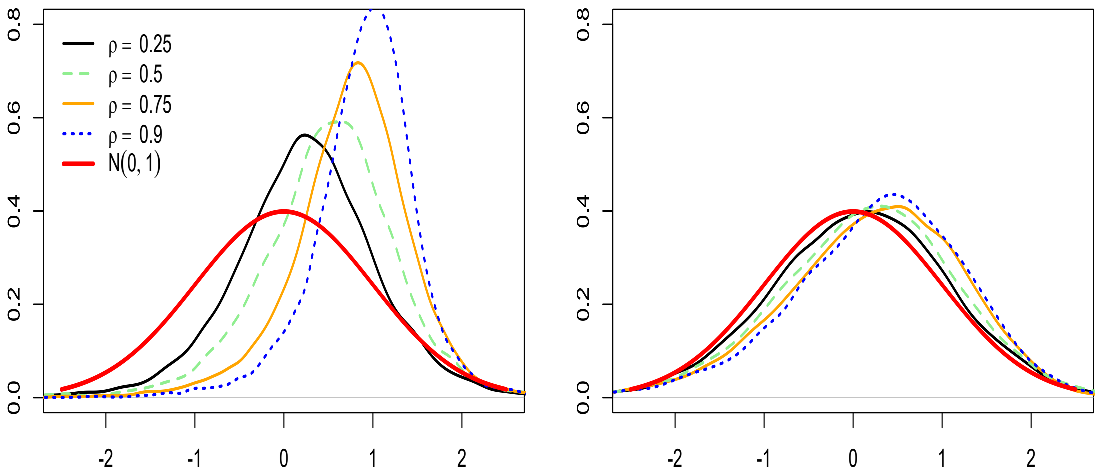


Figure 2.3: Empirical density and ECdf of standardized TSLS estimator under different degrees of identification and $k = 5$ instruments.



(a) $\mu^2/k = 1$ (b) $\mu^2/k = 20$

Figure 2.4: Empirical density of standardized TSLS estimator under weak and strong instruments, with different levels of endogeneity ($k = 5$).

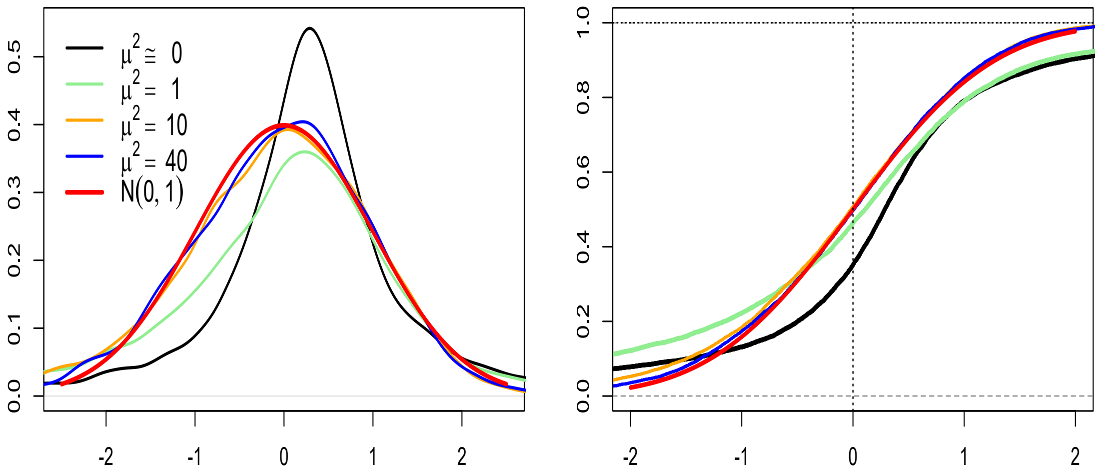
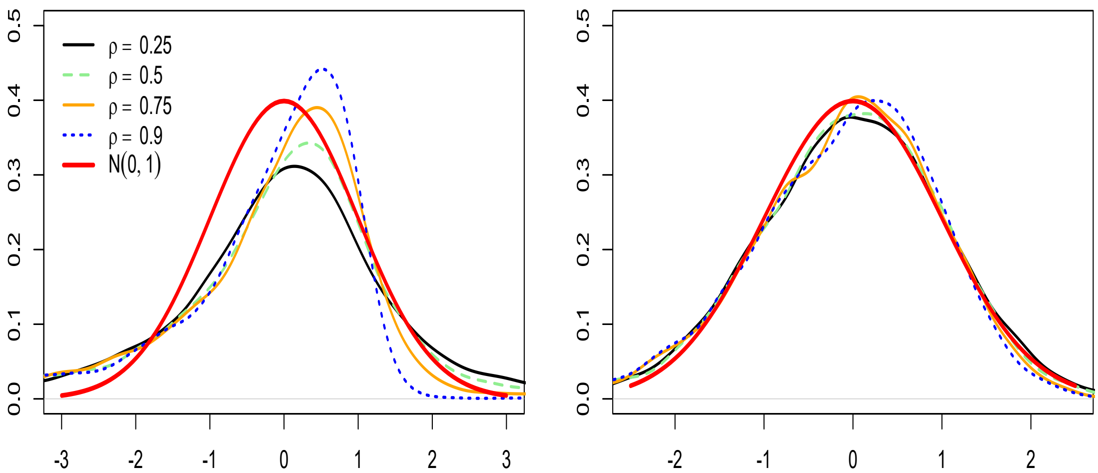


Figure 2.5: Empirical density and ECdf of standardized LIML estimator under different degrees of endogeneity and $k = 5$,



(a) $\mu^2/k = 1$ (b) $\mu^2/k = 20$

Figure 2.6: Empirical density of standardized LIML estimator under weak and strong instruments with different levels of identification ($k = 5$).

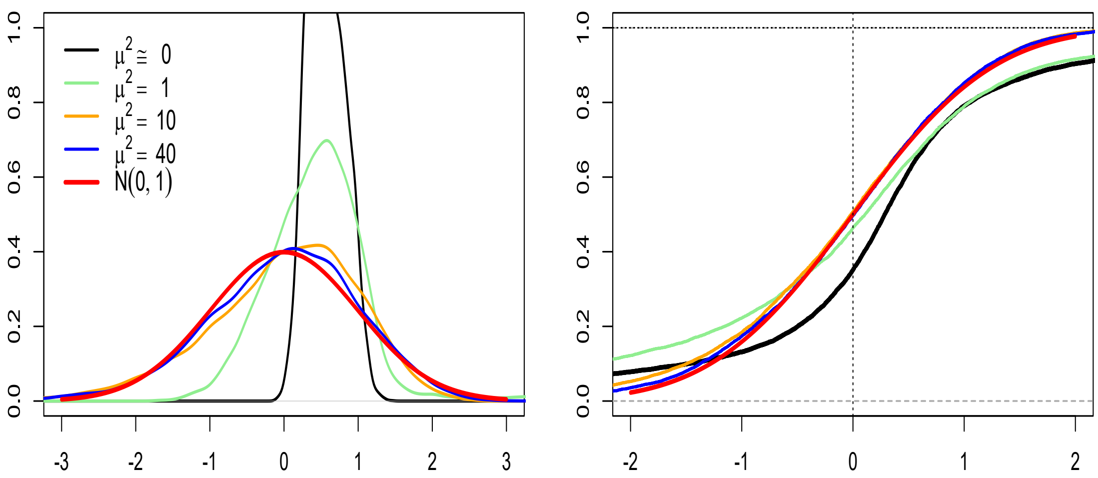
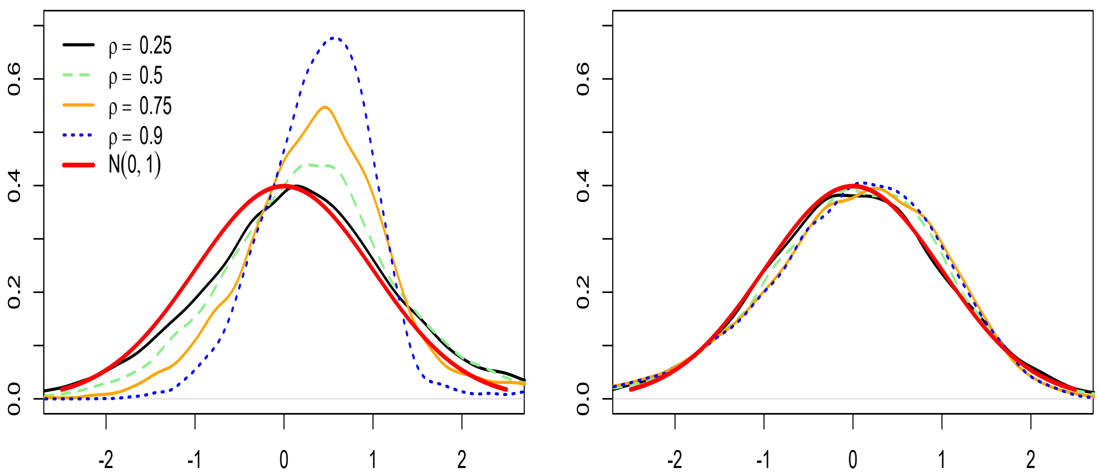


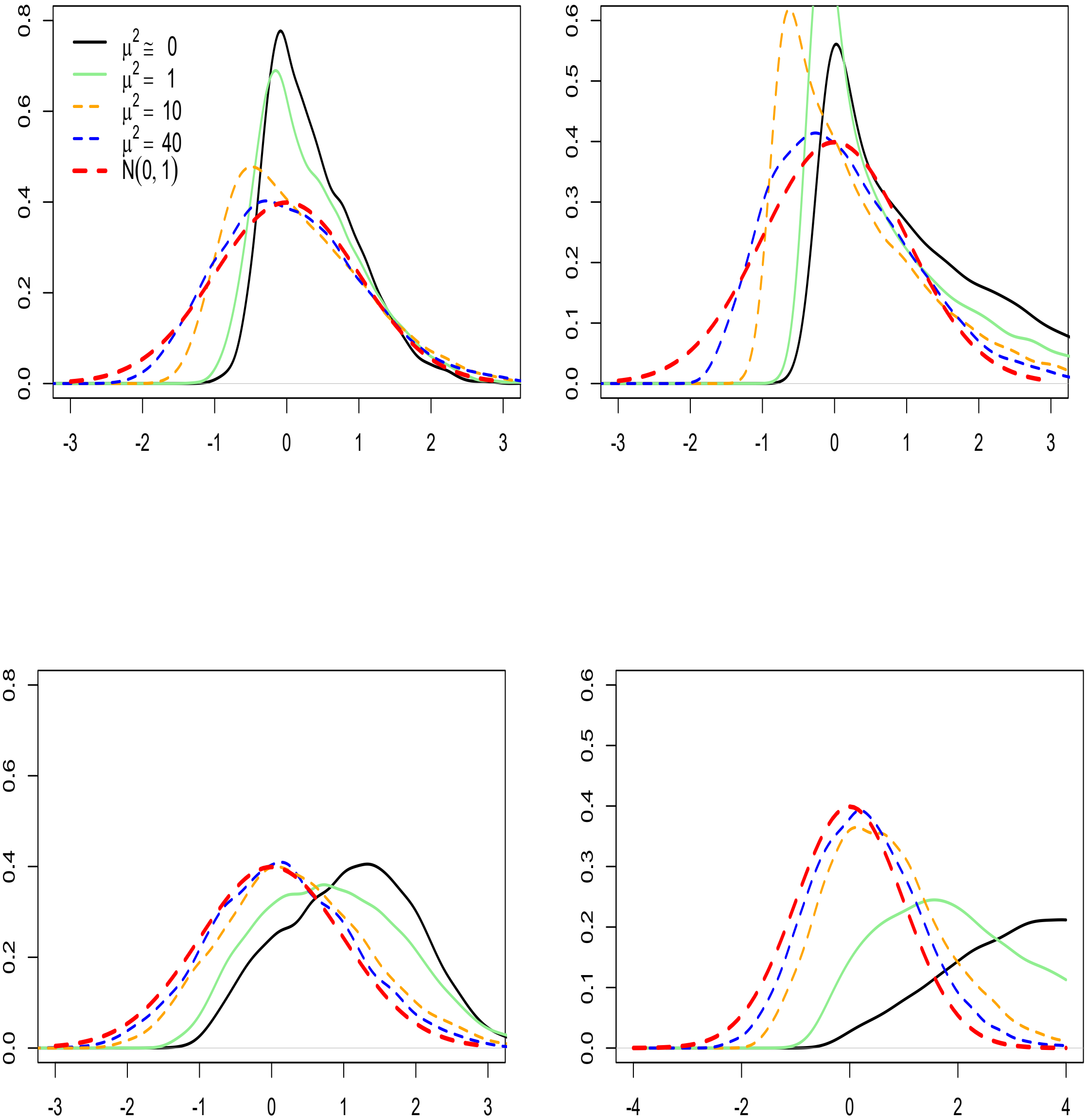
Figure 2.7: Empirical density and ECdf of standardized Fuller estimator under different degrees of identification and $k = 5$ instruments.



(a) $\mu^2/k = 1$

(b) $\mu^2/k = 20$

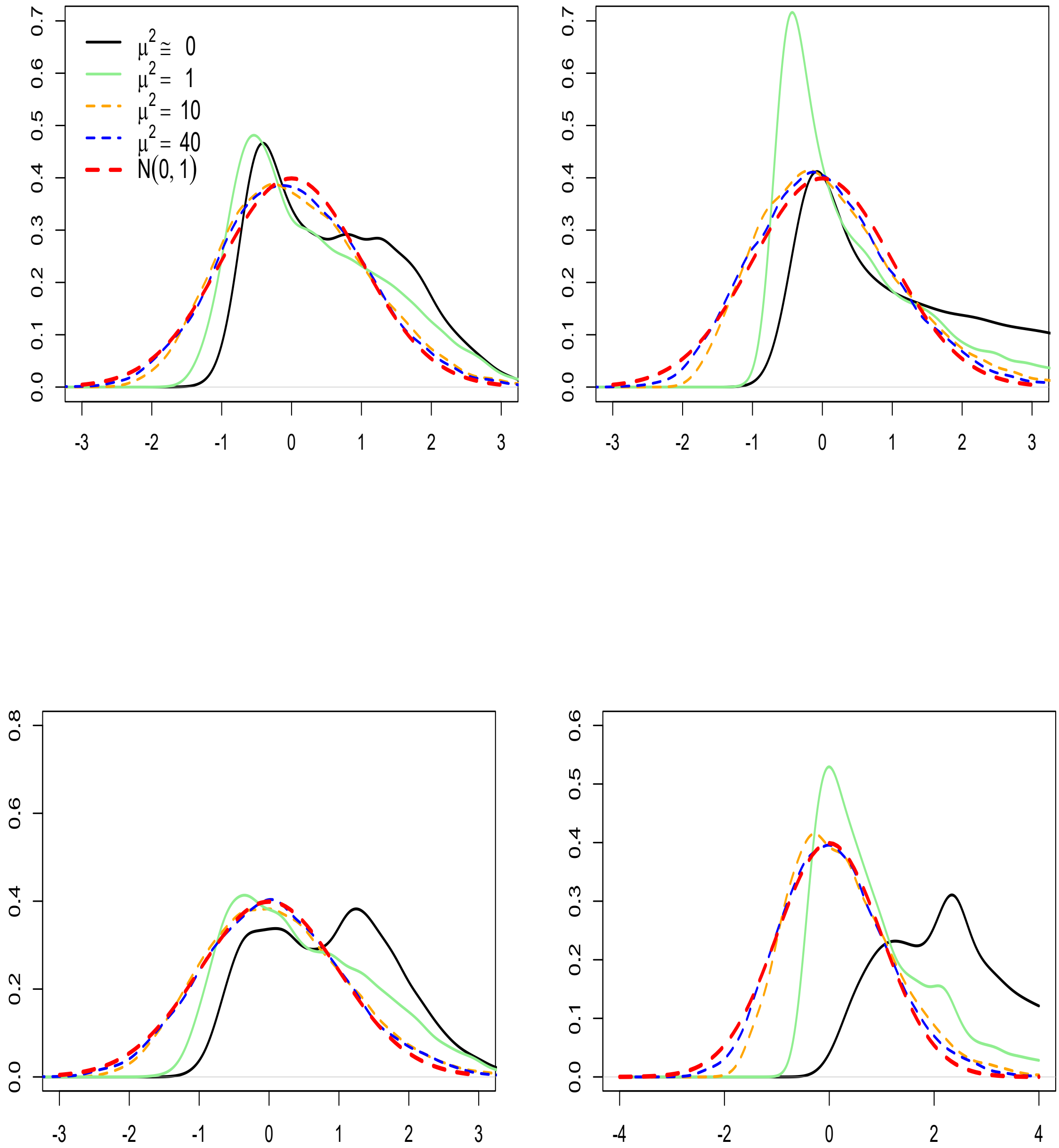
Figure 2.8: Empirical density of standardized Fuller estimator under weak and strong instruments with different levels of endogeneity



(a) $\rho = 0.5$

(b) $\rho = 0.9$

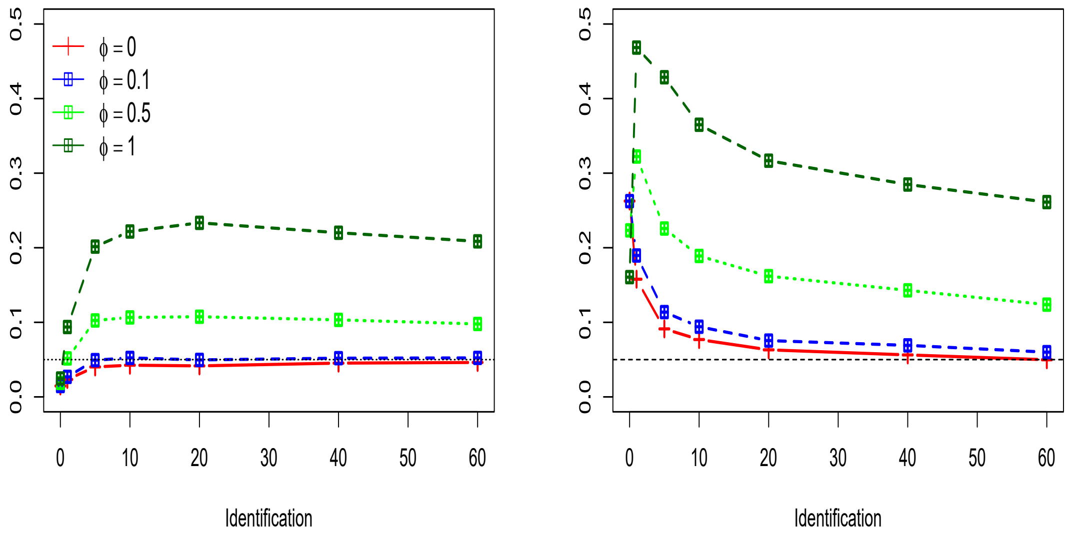
Figure 2.9: Density plot of t–statistic in four identification scenarios under the null hypothesis of $H_0 : \beta = 0$ associated with IV (upper panel, $k = 1$) and TSLS (lower panel, $k = 5$).



(a) $\rho = 0.5$

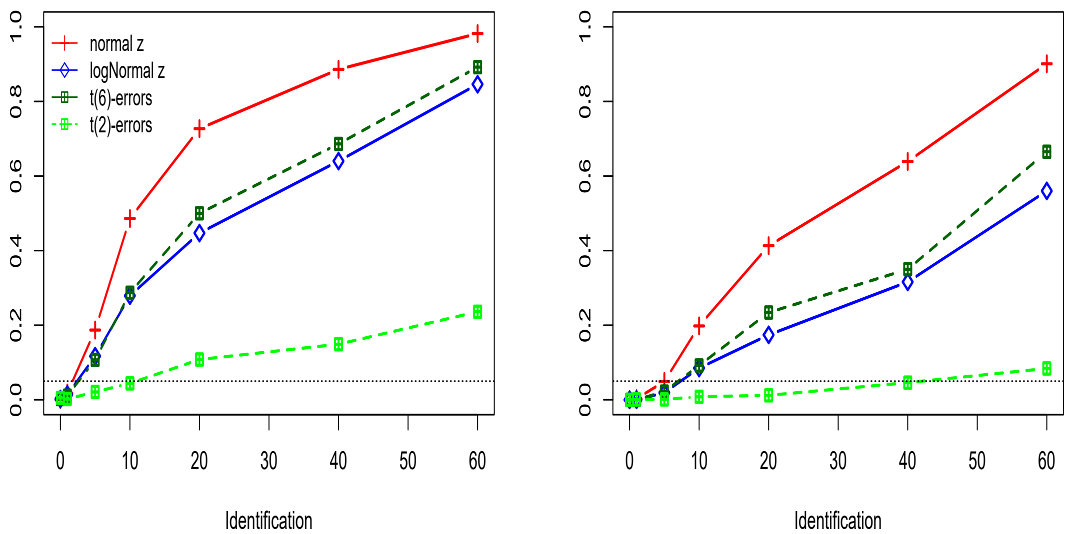
(b) $\rho = 0.9$

Figure 2.11: Density plots of t-statistic in four identification scenarios under the null hypothesis $H_0 = \beta = 0$ associated to LIML (upper panel) and Fuller (lower panel). The number of instruments is $k = 5$.



(a) $\rho = 0.5$ (b) $\rho = 0.9$

Figure 2.12: Rejection frequency of t-test under *weakly endogenous* instruments.



(a) $F > 10$ (b) $F > 16.4$

Figure 2.13: Rejection frequency of F test under different DGPs ($k = 1$).

Design		Performance Indicators					
ρ	μ^2	Median	MAE	RMSE	Coverage	IDR	KS
0.25	0	0.253	1.435	**262.192**	0.998	5.884	—
	1	0.098	0.956	179.199	0.995	3.804	0.075
	5	0.008	0.465	83.178	0.985	1.402	0.062
	10	0.002	0.327	3.896	0.976	0.907	0.045
	20	-0.002	0.229	0.311	0.966	0.608	0.029
	40	0.000	0.160	0.169	0.958	0.422	0.021
	60	0.001	0.131	0.135	0.954	0.340	0.016
	100	-0.001	0.102	0.104	0.951	0.262	0.013
0.50	0	0.499	1.283	**3697.288**	0.984	5.277	—
	1	0.198	0.908	476.588	0.975	3.728	0.095
	5	0.015	0.455	21.601	0.958	1.426	0.079
	10	0.002	0.320	162.044	0.956	0.919	0.061
	20	0.000	0.227	0.282	0.955	0.616	0.044
	40	-0.001	0.161	0.172	0.955	0.424	0.032
	60	0.001	0.131	0.137	0.952	0.341	0.024
	100	0.000	0.101	0.104	0.950	0.262	0.019
0.75	0	0.752	0.982	237.955	0.904	4.099	—
	1	0.286	0.785	130.100	0.908	3.718	0.144
	5	0.017	0.438	49.528	0.922	1.485	0.097
	10	0.001	0.317	**232.568**	0.932	0.947	0.078
	20	-0.001	0.225	0.824	0.942	0.622	0.056
	40	0.000	0.159	0.177	0.947	0.425	0.040
	60	0.000	0.130	0.139	0.949	0.342	0.032
	100	0.001	0.100	0.105	0.949	0.263	0.025
0.90	0	0.901	0.645	147.840	**0.734**	2.686	—
	1	0.302	0.668	415.808	**0.838**	4.057	0.157
	5	0.012	0.426	26.797	0.904	1.551	0.107
	10	0.000	0.313	**19412.531**	0.920	0.967	0.087
	20	-0.001	0.223	0.339	0.934	0.625	0.063
	40	0.000	0.158	0.181	0.943	0.425	0.044
	60	-0.001	0.130	0.141	0.947	0.342	0.037
	100	0.000	0.100	0.106	0.948	0.264	0.030

Table 2.1: Performance of IV estimator under different strength of the instruments and degrees of endogeneity, where $T = 100$.

Design		Performance Indicators					
ρ	μ^2	Median	MAE	RMSE	Coverage	IDR	KS
0.25	0	0.255	1.441	**245.463**	**0.999**	5.942	—
	1	0.103	0.959	64.204	0.996	3.788	0.074
	5	0.007	0.465	16.755	0.986	1.394	0.062
	10	0.001	0.321	1.619	0.978	0.889	0.042
	20	0.000	0.226	1.289	0.969	0.603	0.028
	40	0.000	0.159	0.166	0.961	0.414	0.017
	60	0.000	0.129	0.133	0.957	0.335	0.013
	100	0.000	0.100	0.102	0.955	0.257	0.010
0.50	0	0.500	1.276	549.191	0.985	5.350	—
	1	0.199	0.900	126.712	0.975	3.716	0.098
	5	0.010	0.454	27.572	0.960	1.426	0.079
	10	0.001	0.320	10.407	0.957	0.915	0.061
	20	-0.001	0.225	0.263	0.957	0.605	0.040
	40	0.000	0.158	0.168	0.957	0.415	0.026
	60	0.000	0.130	0.135	0.955	0.339	0.023
	100	0.000	0.101	0.103	0.952	0.260	0.017
0.75	0	0.751	0.979	120.301	0.905	4.081	—
	1	0.279	0.786	63.778	0.909	3.696	0.141
	5	0.013	0.431	**1600.519**	0.924	1.462	0.095
	10	0.000	0.312	2.785	0.935	0.932	0.076
	20	-0.001	0.222	0.282	0.943	0.613	0.053
	40	0.000	0.158	0.173	0.951	0.419	0.037
	60	0.000	0.129	0.137	0.952	0.338	0.031
	100	0.000	0.099	0.104	0.951	0.260	0.024
0.90	0	0.902	0.649	47.548	0.733	2.694	—
	1	0.301	0.666	117.971	**0.838**	4.095	0.156
	5	0.014	0.420	32.067	**0.907**	1.525	0.106
	10	0.002	0.309	7.622	0.921	0.952	0.084
	20	0.000	0.221	0.304	0.933	0.621	0.060
	40	-0.001	0.158	0.177	0.946	0.420	0.043
	60	0.001	0.128	0.138	0.949	0.340	0.035
	100	0.000	0.100	0.104	0.950	0.260	0.026

Table 2.2: Performance of IV estimator under different strength of the instruments and degrees of endogeneity, where $T = 1000$.

Design		Performance Indicators					
ρ	μ^2/k	Median	MAE	RMSE	Coverage	IDR	KS
0.25	0	0.252	0.635	0.982	0.989	1.832	—
	1	0.113	0.440	0.640	0.978	1.244	0.123
	5	0.033	0.244	0.270	0.963	0.649	0.057
	10	0.017	0.177	0.187	0.958	0.464	0.039
	20	0.008	0.128	0.132	0.953	0.331	0.025
	40	0.005	0.091	0.092	0.952	0.234	0.021
	60	0.002	0.074	0.075	0.951	0.191	0.014
	100	0.002	0.058	0.058	0.950	0.148	0.012
0.50	0	0.498	0.567	1.003	0.917	1.639	—
	1	0.230	0.410	0.633	0.910	1.164	0.210
	5	0.062	0.233	0.271	0.931	0.628	0.103
	10	0.033	0.174	0.188	0.941	0.459	0.076
	20	0.016	0.126	0.131	0.946	0.329	0.052
	40	0.008	0.090	0.092	0.947	0.234	0.034
	60	0.005	0.074	0.075	0.947	0.192	0.031
	100	0.003	0.057	0.058	0.949	0.148	0.024
0.75	0	0.749	0.432	1.012	**0.638**	1.254	—
	1	0.341	0.349	0.632	0.752	1.007	0.311
	5	0.093	0.221	0.271	0.884	0.603	0.154
	10	0.048	0.168	0.186	0.914	0.448	0.110
	20	0.025	0.124	0.131	0.932	0.325	0.077
	40	0.013	0.090	0.093	0.940	0.232	0.056
	60	0.008	0.074	0.075	0.945	0.190	0.045
	100	0.005	0.058	0.058	0.946	0.148	0.036
0.90	0	0.899	0.285	1.007	**0.305**	0.823	—
	1	0.404	0.291	0.631	0.617	0.862	0.380
	5	0.112	0.209	0.269	0.851	0.579	0.188
	10	0.059	0.162	0.187	0.897	0.438	0.134
	20	0.029	0.122	0.130	0.924	0.320	0.092
	40	0.015	0.088	0.092	0.936	0.231	0.068
	60	0.010	0.074	0.076	0.939	0.191	0.054
	100	0.006	0.057	0.058	0.943	0.148	0.043

Table 2.3: Performance of TSLS estimator under different strength of the instruments and degrees of endogeneity, where $T = 250$ and $k = 3$.

Design		Performance Indicators					
ρ	μ^2/k	Median	MAE	RMSE	Coverage	IDR	KS
0.25	0	0.249	0.468	0.611	0.972	1.279	—
	1	0.118	0.328	0.394	0.959	0.884	0.157
	5	0.035	0.185	0.197	0.953	0.483	0.078
	10	0.019	0.136	0.140	0.952	0.354	0.055
	20	0.010	0.098	0.100	0.949	0.253	0.041
	40	0.005	0.070	0.071	0.950	0.181	0.030
	60	0.003	0.058	0.058	0.950	0.148	0.023
	100	0.002	0.045	0.045	0.950	0.115	0.018
0.50	0	0.503	0.421	0.709	0.812	1.150	—
	1	0.238	0.303	0.418	0.853	0.823	0.275
	5	0.070	0.179	0.198	0.916	0.469	0.151
	10	0.038	0.133	0.141	0.931	0.346	0.110
	20	0.019	0.097	0.100	0.941	0.251	0.078
	40	0.009	0.069	0.071	0.946	0.179	0.055
	60	0.007	0.057	0.058	0.947	0.148	0.046
	100	0.004	0.045	0.045	0.948	0.114	0.035
0.75	0	0.750	0.319	0.841	**0.406**	0.873	—
	1	0.357	0.257	0.459	0.648	0.704	0.408
	5	0.108	0.166	0.201	0.854	0.441	0.232
	10	0.058	0.127	0.142	0.900	0.336	0.170
	20	0.029	0.095	0.100	0.924	0.247	0.119
	40	0.015	0.069	0.071	0.937	0.177	0.087
	60	0.010	0.057	0.058	0.942	0.146	0.067
	100	0.006	0.044	0.045	0.946	0.113	0.054
0.90	0	0.899	0.210	0.935	**0.115**	0.577	—
	1	0.422	0.214	0.486	0.485	0.594	0.494
	5	0.127	0.157	0.203	0.815	0.420	0.274
	10	0.068	0.124	0.143	0.876	0.327	0.200
	20	0.035	0.093	0.100	0.913	0.242	0.146
	40	0.017	0.068	0.071	0.930	0.176	0.099
	60	0.012	0.057	0.058	0.936	0.146	0.082
	100	0.007	0.044	0.045	0.943	0.113	0.064

Table 2.4: Performance of TSLS estimator under different strength of the instruments and degree of endogeneity, $T = 1000$ and $k = 5$.

Design		Performance Indicators					
ρ	μ^2	Median	MAE	RMSE	Coverage	IDR	KS
0.25	0	0.253	1.429	630.068	0.976	5.973	—
	1	0.025	0.591	140.429	0.962	2.021	0.102
	5	0.000	0.217	0.809	0.954	0.584	0.041
	10	0.000	0.149	0.154	0.951	0.387	0.024
	20	0.000	0.102	0.104	0.951	0.263	0.014
	40	0.000	0.072	0.072	0.951	0.184	0.010
	60	0.000	0.058	0.059	0.949	0.150	0.007
	100	0.000	0.045	0.045	0.950	0.115	0.005
0.50	0	0.497	1.282	**4020.715**	0.868	5.286	—
	1	0.052	0.555	65.076	0.901	1.920	0.104
	5	0.000	0.213	0.369	0.943	0.575	0.050
	10	0.001	0.146	0.154	0.948	0.383	0.031
	20	0.000	0.101	0.105	0.948	0.264	0.021
	40	0.000	0.071	0.072	0.949	0.184	0.013
	60	0.000	0.058	0.058	0.950	0.149	0.011
	100	0.000	0.045	0.045	0.950	0.115	0.008
0.75	0	0.750	0.983	204.125	**0.590**	4.074	—
	1	0.050	0.493	138.579	0.841	1.776	0.111
	5	0.000	0.207	0.260	0.927	0.564	0.056
	10	0.000	0.144	0.155	0.942	0.380	0.037
	20	0.000	0.101	0.104	0.946	0.262	0.025
	40	0.000	0.071	0.073	0.948	0.185	0.019
	60	0.000	0.058	0.059	0.949	0.149	0.014
	100	0.000	0.045	0.045	0.949	0.115	0.011
0.90	0	0.901	0.645	112.420	**0.251**	2.654	—
	1	0.026	0.443	45.714	0.823	1.644	0.113
	5	0.000	0.201	0.253	0.920	0.554	0.057
	10	0.001	0.142	0.157	0.936	0.380	0.042
	20	0.000	0.100	0.105	0.945	0.262	0.029
	40	0.000	0.071	0.072	0.948	0.183	0.020
	60	0.000	0.058	0.059	0.949	0.149	0.016
	100	0.000	0.045	0.045	0.949	0.116	0.013

Table 2.5: Performance of LIML estimator under different strength of the instruments and degree of endogeneity, where $T = 1000$ and $k = 5$.

Design		Performance Indicators					
ρ	μ^2	Median	MAE	RMSE	Coverage	IDR	KS
0.25	0	0.253	0.694	0.688	0.981	1.677	—
	1	0.066	0.461	0.507	0.965	1.256	0.059
	5	0.009	0.209	0.228	0.952	0.554	0.024
	10	0.006	0.144	0.149	0.951	0.375	0.017
	20	0.002	0.101	0.103	0.950	0.260	0.009
	40	0.001	0.071	0.072	0.950	0.183	0.009
	60	0.001	0.058	0.058	0.950	0.149	0.005
	100	0.001	0.045	0.045	0.950	0.115	0.006
0.50	0	0.499	0.619	0.761	0.899	1.498	—
	1	0.133	0.415	0.486	0.915	1.138	0.128
	5	0.022	0.200	0.220	0.948	0.534	0.043
	10	0.010	0.142	0.148	0.951	0.370	0.028
	20	0.005	0.100	0.103	0.950	0.260	0.021
	40	0.002	0.071	0.072	0.951	0.182	0.013
	60	0.002	0.058	0.058	0.950	0.148	0.012
	100	0.001	0.045	0.045	0.949	0.115	0.010
0.75	0	0.752	0.474	0.869	**0.699**	1.142	—
	1	0.181	0.330	0.435	**0.878**	0.888	0.203
	5	0.029	0.189	0.211	0.940	0.510	0.061
	10	0.015	0.137	0.146	0.949	0.361	0.044
	20	0.007	0.099	0.102	0.950	0.257	0.031
	40	0.004	0.070	0.071	0.950	0.181	0.022
	60	0.002	0.058	0.058	0.949	0.148	0.016
	100	0.001	0.044	0.045	0.949	0.115	0.012
0.90	0	0.899	0.312	0.945	**0.497**	0.753	—
	1	0.198	0.259	0.370	**0.885**	0.663	0.267
	5	0.037	0.180	0.206	0.935	0.492	0.077
	10	0.017	0.135	0.145	0.947	0.356	0.050
	20	0.009	0.098	0.102	0.949	0.256	0.035
	40	0.005	0.070	0.071	0.950	0.180	0.026
	60	0.003	0.057	0.058	0.950	0.147	0.024
	100	0.001	0.044	0.045	0.949	0.115	0.014

Table 2.6: Performance of Fuller estimator under different strength of the instruments and degrees of endogeneity, where $T = 1000$ and $k = 5$.

Design			Performance Indicators					
ρ	μ^2	ϕ/π	Median	MAE	RMSE	Coverage	IDR	KS
0.25	0	—	0.240	1.437	**1852.240**	0.998	5.840	—
	1	0.100	0.179	0.971	24.928	0.996	3.833	0.088
	5	0.045	0.058	0.449	4.693	0.984	1.340	0.054
	10	0.032	0.027	0.320	0.660	0.971	0.896	0.037
	20	0.022	0.021	0.229	0.256	0.961	0.610	0.040
	40	0.016	0.014	0.161	0.167	0.958	0.418	0.041
	60	0.013	0.014	0.130	0.134	0.953	0.337	0.047
	100	0.010	0.010	0.100	0.102	0.948	0.256	0.046
0.50	0	—	0.473	1.282	45.062	0.984	5.277	—
	1	0.100	0.265	0.871	55.301	0.972	3.735	0.133
	5	0.045	0.058	0.443	8.066	0.948	1.392	0.069
	10	0.032	0.031	0.317	2.645	0.945	0.898	0.047
	20	0.022	0.021	0.218	0.257	0.948	0.600	0.040
	40	0.016	0.015	0.157	0.166	0.950	0.411	0.042
	60	0.013	0.015	0.129	0.133	0.951	0.334	0.049
	100	0.010	0.009	0.101	0.102	0.949	0.260	0.040
0.75	0	—	0.743	0.993	89.590	0.906	4.138	—
	1	0.100	0.347	0.752	71.611	0.899	3.425	0.172
	5	0.045	0.061	0.413	**255.067**	0.910	1.396	0.083
	10	0.032	0.032	0.308	7.527	0.920	0.924	0.066
	20	0.022	0.022	0.218	0.266	0.930	0.607	0.043
	40	0.016	0.014	0.156	0.168	0.938	0.416	0.036
	60	0.013	0.011	0.129	0.137	0.939	0.336	0.036
	100	0.010	0.008	0.099	0.101	0.947	0.255	0.033
0.90	0	—	0.898	0.652	123.809	**0.734**	2.657	—
	1	0.100	0.383	0.641	39.221	**0.807**	3.604	**0.194**
	5	0.045	0.051	0.414	**262.125**	0.886	1.496	0.096
	10	0.032	0.029	0.304	0.943	0.906	0.950	0.077
	20	0.022	0.024	0.216	0.282	0.919	0.608	0.047
	40	0.016	0.015	0.157	0.172	0.931	0.414	0.043
	60	0.013	0.013	0.125	0.135	0.941	0.336	0.045
	100	0.010	0.010	0.099	0.105	0.938	0.260	0.040

Table 2.7: Performance of IV estimator with a single weakly endogenous instrument, $\phi = 0.1$.

Design			Performance Indicators					
ρ	μ^2	ϕ/π	Median	MAE	RMSE	Coverage	IDR	KS
0.25	0	—	0.222	2.234	80.652	0.996	9.050	—
	1	1.000	0.721	1.051	40.905	0.977	4.523	0.298
	5	0.447	0.440	0.452	5.847	0.910	1.351	0.385
	10	0.316	0.315	0.311	0.773	0.866	0.856	0.386
	20	0.224	0.222	0.217	0.325	0.828	0.581	0.392
	40	0.158	0.157	0.152	0.223	0.816	0.396	0.396
	60	0.129	0.128	0.126	0.180	0.821	0.322	0.384
	100	0.100	0.097	0.101	0.140	0.816	0.258	0.375
0.5	0	—	0.495	2.179	118.083	0.975	8.788	—
	1	1.000	0.775	0.906	23.209	0.908	3.659	0.341
	5	0.447	0.451	0.407	3.311	0.797	1.226	0.401
	10	0.316	0.317	0.288	0.858	0.763	0.791	0.393
	20	0.224	0.228	0.205	0.312	0.757	0.546	0.403
	40	0.158	0.157	0.147	0.214	0.776	0.386	0.390
	60	0.129	0.128	0.121	0.175	0.783	0.319	0.390
	100	0.100	0.100	0.095	0.137	0.794	0.248	0.390
0.75	0	—	0.785	2.084	449.246	0.908	8.041	—
	1	1.000	0.897	0.667	136.609	0.736	2.766	0.430
	5	0.447	0.453	0.332	17.775	0.652	1.026	0.425
	10	0.316	0.317	0.250	0.892	0.676	0.713	0.409
	20	0.224	0.225	0.189	0.298	0.702	0.516	0.397
	40	0.158	0.160	0.143	0.210	0.728	0.371	0.399
	60	0.129	0.133	0.115	0.172	0.751	0.312	0.400
	100	0.100	0.100	0.095	0.134	0.766	0.242	0.389
0.90	0	—	0.825	2.115	863.908	0.844	7.782	—
	1	1.000	0.960	0.426	35.205	**0.526**	1.698	0.525
	5	0.447	0.458	0.274	42.674	0.558	0.914	0.442
	10	0.316	0.320	0.226	2.208	0.627	0.685	0.413
	20	0.224	0.224	0.180	0.307	0.676	0.495	0.403
	40	0.158	0.159	0.138	0.205	0.707	0.362	0.396
	60	0.129	0.130	0.112	0.167	0.734	0.299	0.400
	100	0.100	0.101	0.092	0.132	0.749	0.236	0.393

Table 2.8: Performance of IV estimator under a weak and endogenous instrument, $\phi = 1$

$k = 3$

ρ	μ^2/k	$\bar{\tilde{\beta}}_T$	$Me(\tilde{\beta}_T)$	$V(\tilde{\beta}_T)$	$Skew(\tilde{\beta}_T)$	$K(\tilde{\beta}_T)$	IQR	KS
0.5	1	0.350	0.396	1.133	-4.242	281.188	0.958	0.209
	5	0.143	0.247	1.087	-1.099	11.265	1.230	0.107
	10	0.090	0.171	1.039	-0.656	4.607	1.285	0.072
	20	0.070	0.127	1.009	-0.408	3.538	1.310	0.053
	40	0.047	0.088	1.005	-0.271	3.204	1.330	0.036
	60	0.039	0.074	1.010	-0.228	3.176	1.343	0.030
0.9	1	0.635	0.695	0.771	0.293	60.200	0.691	0.379
	5	0.253	0.433	1.026	-2.369	32.664	1.118	0.187
	10	0.167	0.310	1.012	-1.234	7.722	1.226	0.130
	20	0.119	0.230	1.019	-0.785	4.464	1.287	0.097
	40	0.083	0.163	1.003	-0.503	3.530	1.318	0.066
	60	0.073	0.132	0.996	-0.417	3.388	1.321	0.055

$k = 5$

ρ	μ^2/k	$\bar{\tilde{\beta}}_T$	$Me(\tilde{\beta}_T)$	$V(\tilde{\beta}_T)$	$Skew(\tilde{\beta}_T)$	$K(\tilde{\beta}_T)$	IQR	KS
0.5	1	0.502	0.533	0.634	-0.274	6.445	0.910	0.276
	5	0.293	0.360	0.893	-0.538	4.110	1.200	0.155
	10	0.204	0.261	0.954	-0.399	3.580	1.276	0.108
	20	0.152	0.201	0.983	-0.305	3.302	1.318	0.081
	40	0.102	0.139	0.988	-0.230	3.163	1.330	0.057
	60	0.088	0.113	0.995	-0.167	3.090	1.338	0.047
0.9	1.	0.898	0.944	0.385	-1.809	82.622	0.649	0.495
	5	0.518	0.638	0.765	-1.081	5.916	1.073	0.277
	10	0.374	0.477	0.885	-0.793	4.503	1.198	0.199
	20	0.268	0.349	0.942	-0.547	3.639	1.272	0.142
	40	0.186	0.248	0.973	-0.385	3.305	1.311	0.100
	60	0.159	0.212	0.985	-0.313	3.210	1.322	0.086

Table 2.9: Standardized TSLS estimator with $T = 1000$ and $k = 3, 5$.

$DF = 12$

ρ	μ^2/k	$\bar{\tilde{\beta}}_T$	$Me(\tilde{\beta}_T)$	$V(\tilde{\beta}_T)$	$Skew(\tilde{\beta}_T)$	$K(\tilde{\beta}_T)$	IQR	KS
0.5	1	0.548	0.583	0.724	-0.550	7.688	0.953	0.287
	5	0.346	0.421	1.058	-0.476	3.870	1.310	0.169
	10	0.240	0.301	1.130	-0.427	3.538	1.361	0.120
	20	0.172	0.222	1.187	-0.309	3.279	1.447	0.091
	40	0.135	0.174	1.167	-0.235	3.135	1.433	0.073
	60	0.121	0.148	1.171	-0.196	3.130	1.426	0.065
0.9	1	1.011	1.046	0.402	-0.108	9.252	0.641	0.533
	5	0.613	0.740	0.879	-1.139	5.814	1.134	0.313
	10	0.439	0.561	1.012	-0.875	4.949	1.283	0.222
	20	0.319	0.431	1.140	-0.566	3.523	1.404	0.169
	40	0.230	0.323	1.182	-0.466	3.567	1.429	0.128
	60	0.182	0.245	1.207	-0.382	3.402	1.476	0.099

$DF = 2$

ρ	μ^2/k	$\bar{\tilde{\beta}}_T$	$Me(\tilde{\beta}_T)$	$V(\tilde{\beta}_T)$	$Skew(\tilde{\beta}_T)$	$K(\tilde{\beta}_T)$	IQR	KS
0.5	1	0.613	0.625	2.197	-3.849	129.400	1.259	0.263
	5	0.541	0.620	4.346	0.302	29.120	1.909	0.237
	10	0.391	0.492	5.361	2.009	110.404	2.081	0.209
	20	0.247	0.348	5.392	0.118	35.057	2.188	0.184
	40	0.231	0.268	5.163	-0.596	57.637	2.299	0.179
	60	0.216	0.230	5.082	2.595	68.101	2.292	0.168
0.9	1	1.088	1.098	0.971	-2.000	60.918	0.814	0.511
	5	0.976	1.108	3.593	-3.782	129.565	1.536	0.400
	10	0.730	0.885	3.685	-1.917	50.434	1.796	0.315
	20	0.557	0.718	4.451	-0.869	20.311	2.048	0.267
	40	0.388	0.507	4.393	0.782	25.571	2.187	0.218
	60	0.324	0.394	4.483	-0.111	23.401	2.193	0.197

Table 2.10: Standardized TSLS estimator under t(12) and t(2) disturbances, with $T = 1000$ and $k = 3$.

LIML

ρ	μ^2/k	$\bar{\tilde{\beta}}_T$	$Me(\tilde{\beta}_T)$	$V(\tilde{\beta}_T)$	$Skew(\tilde{\beta}_T)$	$K(\tilde{\beta}_T)$	IQR	KS
0.5	1	-0.363	0.114	**1617.946**	6.475	2262.168	1.749	0.105
	5	-0.136	-0.007	1.626	1.108	64.444	1.458	0.051
	10	-0.096	-0.008	1.175	-0.549	3.995	1.399	0.035
	20	-0.070	-0.009	1.094	-0.345	3.369	1.355	0.024
	40	-0.024	0.018	1.051	-0.263	3.132	1.352	0.018
	60	-0.028	0.004	1.048	-0.153	3.039	1.374	0.015
0.9	1	-0.608	0.023	**3995.979**	-7.163	2327.911	1.518	0.118
	5	-0.225	-0.005	1.548	-1.986	14.075	1.419	0.060
	10	-0.136	0.011	1.207	-0.964	4.977	1.368	0.042
	20	-0.092	0.016	1.097	-0.623	3.843	1.346	0.030
	40	-0.072	-0.016	1.062	-0.388	3.395	1.369	0.022
	60	-0.043	0.008	1.050	-0.286	3.218	1.365	0.019

Fuller($\underline{c} = 1$)

ρ	μ^2/k	$\bar{\tilde{\beta}}_T$	$Me(\tilde{\beta}_T)$	$V(\tilde{\beta}_T)$	$Skew(\tilde{\beta}_T)$	$K(\tilde{\beta}_T)$	IQR	KS
0.5	1	0.317	0.296	1.050	0.226	3.842	1.232	0.130
	5	-0.028	0.080	1.248	-0.819	5.335	1.362	0.033
	10	-0.006	0.072	1.078	-0.499	3.788	1.350	0.032
	20	0.027	0.072	1.055	-0.359	3.498	1.361	0.031
	40	-0.012	0.021	1.024	-0.179	3.116	1.351	0.010
	60	-0.020	0.001	1.017	-0.172	3.121	1.342	0.010
0.9	1	0.432	0.441	0.468	1.144	8.589	0.788	0.266
	5	0.010	0.160	1.087	-1.193	5.980	1.266	0.071
	10	0.005	0.118	1.039	-0.812	4.424	1.306	0.049
	20	-0.005	0.082	1.008	-0.522	3.572	1.330	0.035
	40	-0.002	0.071	1.028	-0.444	3.421	1.344	0.030
	60	0.004	0.050	1.020	-0.264	3.003	1.364	0.022

Table 2.11: Standardized $\kappa-$class estimators with $T = 1000$ and $k = 5$.

T=100

ρ	μ^2/k	OLS	TSLS	LIML	Fuller(1)	Fuller(4)
	0	1.000	0.181	0.097	0.134	**0.241**
	1	1.000	0.143	0.082	0.095	0.150
	5	0.998	0.082	0.052	0.058	0.081
0.5	10	0.990	0.065	0.048	0.049	0.064
	20	0.954	0.058	0.047	0.048	0.054
	40	0.831	0.056	0.051	0.051	0.054
	60	0.712	0.052	0.048	0.048	0.050
	0	1.000	0.882	0.499	0.768	**0.999**
	1	1.000	0.510	0.111	0.177	0.581
	5	1.000	0.185	0.062	0.080	0.159
0.9	10	1.000	**0.123**	0.052	0.064	0.108
	20	1.000	**0.090**	0.052	0.058	0.083
	40	1.000	0.070	0.047	0.050	0.065
	60	0.998	0.065	0.052	0.055	0.063

T=1000

ρ	μ^2/k	OLS	TSLS	LIML	Fuller(1)	Fuller(4)
	0	1.000	0.178	0.097	0.132	0.243
	1	1.000	0.146	0.085	0.100	0.153
	5	1.000	0.087	0.053	0.059	0.080
0.5	10	1.000	0.068	0.050	0.054	0.066
	20	1.000	0.062	0.050	0.053	0.062
	40	1.000	0.056	0.052	0.052	0.054
	60	1.000	0.057	0.053	0.053	0.056
	0	1.000	0.888	0.502	0.774	1.000
	1	1.000	0.512	0.113	0.182	0.584
	5	1.000	0.185	0.062	0.078	0.157
0.9	10	1.000	0.128	0.057	0.068	0.118
	20	1.000	0.095	0.054	0.059	0.086
	40	1.000	0.071	0.053	0.057	0.066
	60	1.000	0.063	0.047	0.050	0.059

Table 2.12: Rejection frequencies (empirical size) of t–test associated to κ–class estimators under weak instruments and (jointly) normal disturbances.

$\phi = 0.1$

ρ	μ^2/k	OLS	TSLS	LIML	Fuller(1)	Fuller(4)
	0	1.000	0.085	0.058	0.084	0.179
	1	1.000	0.094	0.065	0.080	0.140
	5	0.999	0.071	0.053	0.062	0.092
0.5	10	0.997	0.065	0.050	0.055	0.075
	20	0.987	0.055	0.050	0.052	0.064
	40	0.944	0.056	0.051	0.052	0.059
	60	0.866	0.051	0.050	0.050	0.054
	0	1.000	0.884	0.494	0.763	0.999
	1	1.000	0.520	0.120	0.187	0.595
	5	1.000	0.201	0.068	0.085	0.169
0.9	10	1.000	**0.136**	0.059	0.070	0.118
	20	1.000	**0.088**	0.052	0.057	0.080
	40	1.000	0.078	0.055	0.059	0.076
	60	0.998	0.066	0.050	0.053	0.063

$\phi = 1$

ρ	μ^2/k	OLS	TSLS	LIML	Fuller(1)	Fuller(4)
	0	0.999	0.197	0.100	0.134	0.252
	1	1.000	0.235	0.144	0.163	0.237
	5	0.999	0.173	0.107	0.117	0.156
0.5	10	0.996	0.133	0.088	0.096	0.124
	20	0.980	0.120	0.090	0.096	0.114
	40	0.912	0.098	0.077	0.080	0.093
	60	0.834	0.104	0.088	0.092	0.101
	0	1.000	0.606	0.246	0.598	0.988
	1	1.000	0.577	0.219	0.386	0.926
	5	1.000	0.335	0.159	0.202	0.392
0.9	10	1.000	0.255	0.143	0.172	0.289
	20	1.000	0.197	0.133	0.152	0.222
	40	1.000	0.167	0.119	0.133	0.179
	60	1.000	0.142	0.111	0.121	0.158

Table 2.13: Rejection frequencies (empirical size, $T = 100$) of t–test associated to $\kappa-$class estimators ($k = 5$) under weak and weakly endogenous (invalid) instruments.

Design			First Stage($k=1$)						
ρ	μ^2	Bias	MC–bias	$\overline{R}_f^2$	$Me(R_f^2)$	$Me(F)$	$\hat{\mu}^2$	$F>10$	$F>16.4$
	0		2.900	0.001	0.000	0.448	0.000	0.001	0.000
	1	-0.500	-0.234	0.002	0.001	1.125	1.033	0.015	0.001
	5	-0.100	-0.068	0.006	0.005	4.866	4.933	0.176	0.034
0.5	10	-0.050	-0.080	0.011	0.010	9.935	10.008	0.496	0.186
	20	-0.025	-0.031	0.020	0.020	19.899	19.938	0.903	0.659
	40	-0.012	-0.015	0.039	0.038	39.827	39.950	0.999	0.988
	60	-0.008	-0.008	0.057	0.056	59.674	59.959	1.000	1.000
	0		0.644	0.001	0.000	0.454	0.003	0.001	0.000
	1	-0.900	0.534	0.002	0.001	1.131	1.040	0.017	0.001
	5	-0.180	-0.421	0.006	0.005	4.944	4.921	0.173	0.035
0.9	10	-0.090	-0.353	0.011	0.010	10.018	10.017	0.501	0.188
	20	-0.045	-0.054	0.021	0.020	19.967	20.096	0.904	0.664
	40	-0.022	-0.022	0.039	0.039	40.063	40.168	0.999	0.987
	60	-0.015	-0.016	0.057	0.057	59.886	60.115	1.000	1.000

			$k=5$						
ρ	μ^2	Bias	MC–bias	$\overline{R}_f^2$	$Me(R_f^2)$	$Me(F)$	$\hat{\mu}^2/k$	$F>10$	$F>16.4$
	0		0.500	0.005	0.004	0.863	0.000	0.000	0.000
	1	0.300	0.222	0.010	0.009	1.837	1.000	0.000	0.000
	5	0.060	0.059	0.029	0.028	5.813	5.003	0.043	0.024
0.5	10	0.030	0.031	0.052	0.051	10.743	9.999	0.601	0.488
	20	0.015	0.014	0.095	0.095	20.803	20.017	0.999	0.998
	40	0.008	0.008	0.171	0.170	40.837	40.131	1.000	1.000
	60	0.005	0.006	0.235	0.235	60.903	60.181	1.000	1.000
	0		0.903	0.005	0.004	0.874	0.002	0.000	0.000
	1	0.540	0.407	0.010	0.009	1.812	0.993	0.000	0.000
	5	0.108	0.103	0.029	0.028	5.788	4.996	0.045	0.024
0.9	10	0.054	0.053	0.052	0.052	10.834	10.056	0.610	0.501
	20	0.027	0.028	0.096	0.095	20.881	20.106	0.999	0.997
	40	0.014	0.012	0.170	0.170	40.775	39.986	1.000	1.000
	60	0.009	0.008	0.234	0.234	60.745	60.132	1.000	1.000

Table 2.14: Bias of IV/TSLS and evidences from first stage with $T=100$ under jointly normal disturbances and two levels of endogeneity.

Design			First Stage (k=1)						
ρ	μ^2	Bias	MC–bias	$\overline{R}_f^2$	$Me(R_f^2)$	$Me(F)$	$\hat{\mu}^2$	$F > 10$	$F > 10.83$
	0		0.271	0.001	0.000	0.449	0.018	0.002	0.001
	1	-0.500	-2.592	0.002	0.001	0.850	0.666	0.010	0.007
	5	-0.100	-0.228	0.004	0.003	3.332	3.318	0.089	0.071
0.5	10	-0.050	-0.093	0.008	0.007	6.661	6.673	0.281	0.240
	20	-0.025	-0.085	0.014	0.013	13.413	13.511	0.689	0.644
	40	-0.012	-0.025	0.027	0.026	26.621	26.736	0.978	0.968
	60	-0.008	-0.014	0.040	0.039	40.259	40.273	0.999	0.998
	0		1.246	0.001	0.000	0.465	0.019	0.002	0.000
	1	-0.900	0.272	0.002	0.001	0.804	0.619	0.008	0.000
	5	-0.180	0.054	0.004	0.003	3.423	3.384	0.090	0.012
0.9	10	-0.090	-0.044	0.008	0.007	6.645	6.672	0.280	0.072
	20	-0.045	-0.049	0.014	0.013	13.194	13.306	0.681	0.344
	40	-0.022	-0.039	0.027	0.026	26.720	26.726	0.975	0.864
	60	-0.015	-0.022	0.040	0.039	40.356	40.342	0.999	0.986

					$k = 5$				
ρ	μ^2	Bias	MC–bias	$\overline{R}_f^2$	$Me(R_f^2)$	$Me(F)$	$\hat{\mu}^2/k$	$F > 10$	$F > 10.83$
	0		0.504	0.005	0.004	0.870	0.003	0.000	0.000
	1	0.300	0.279	0.008	0.008	1.509	0.671	0.000	0.000
	5	0.060	0.082	0.021	0.020	4.118	3.324	0.005	0.002
0.5	10	0.030	0.044	0.037	0.036	7.476	6.715	0.172	0.112
	20	0.015	0.022	0.067	0.066	14.114	13.392	0.905	0.848
	40	0.008	0.012	0.122	0.122	27.584	26.779	1.000	1.000
	60	0.005	0.008	0.171	0.171	40.944	40.195	1.000	1.000
	0		0.897	0.005	0.004	0.875	0.006	0.000	0.000
	1	0.540	0.509	0.008	0.007	1.500	0.670	0.000	0.000
	5	0.108	0.153	0.021	0.021	4.178	3.367	0.005	0.002
0.9	10	0.054	0.082	0.037	0.037	7.545	6.731	0.174	0.107
	20	0.027	0.039	0.067	0.066	14.104	13.369	0.899	0.842
	40	0.014	0.020	0.122	0.122	27.538	26.762	1.000	1.000
	60	0.009	0.013	0.171	0.171	40.882	40.117	1.000	1.000

Table 2.15: Bias of TSLS and evidences from first stage with $T = 100$ and two levels of endogeneity. Disturbances are distributed as a multivariate t(6).

Chapter 3

Bootstrap asymptotics under weak instruments

In this Chapter we investigate the bootstrapped distribution of instrumental variables estimators under weak instrument asymptotics in presence of one endogenous regressor, both from a theoretical point of view and through simulation. Firstly, we consider the just identified case and then models involving several instruments. Under strong instrument asymptotics, i.e. $\boldsymbol{\pi} = \boldsymbol{\pi}_0$, the bootstrapped IV/TSLS estimator converges in distribution to its proper normal limit, matching the true value of the structural parameter β and its (theoretical) asymptotic variance $V(\hat{\beta}_T)$. However, when low relevance issue arises, non-normality of IV/TSLS estimators, discussed throughout Chapter 2, could be reflected in their bootstrap counterpart.

From a theoretical point of view, we propose a new formulation of the bootstrapped IV/TSLS estimator under weak instruments and homoskedastic disturbances, applying unrestricted residual bootstrap both with fixed and resampled instruments. We consider these methods because they are applied both in theoretical and empirical works, being moderately straightforward to implement. Therefore, they are also not too computationally demanding, requiring only parameter estimates and residuals from the unrestricted model. We also consider other bootstrap methods, summarized in Section 2.4, finding similar problems in confidence intervals and p–values when strength of instruments is very low. On one hand we notice that the distribution of bootstrapped estimator, conditionally on the sample, preserves some components of those derived in Stock and Yogo (2005), which mainly concerns its non-normality. This could be intepreted as a pattern of weak identification and should be a desirable feature in practice, in order to recognize weak instruments via bootstrap methods. On the other hand, this limit distribution does not asymptotically mimic its weak instrument asymptotics counterpart. Thus, result of conventional bootstrap inference does no longer hold even if $T \to \infty$ and $B \to \infty$.

Bootstrap methods are generally deemed not valid in IV/TSLS (and other $\kappa-$class) inference when instruments are poorly correlated with the endogenous regressor, and their use is often discouraged in point estimation and hypothesis testing. Neverthless, Zhan (2017) approaches this issue arguing that malfunctions of IV/TSLS bootstrap inference strictlty depends on the asymptotic difference between the concentration parameter μ^2 and its bootstrap counterpart μ^{*2}, occurring when instruments are weak or nearly irrelevant. He also argues that bootstrap distribution of estimators can provide information on the identification strength; malfunctions of conventional inference may be recognized in practice. To summarize, the bootstrap counterpart of concentration parameter, μ^{*2}, is different (greater) than μ^2, although both converges to constants, and the bootstrap sistematically fails to estimate the level of identification strength. For this reason, both the estimated first stage F statistic and its bootstrap counterpart F^* could be different from population F under weak instruments asymptotics. However, in our simulation studies, we found that proportion of bootstrapped $F^* > 10$ is often preserved in the bootstrap world.

As mentioned, we give a different interpretation of the failure regarding bootstrap in linear IV/GMM, related to a *randomness* of bootstrapped distributions under weak instrument asymptotics, conditionally on the data D_T. Consider a specific (test) statistic or parameter estimator τ, converging in distribution to a non–dengeerate random variable τ_∞, and its bootstrap counterpart τ^*. Although validity requires that cumulative distribution of τ^*, denoted with G_T^*, converges Cdf of τ_∞ (i.e. G_∞), conditionally on the original data, there are cases in which the bootstrap distribution of τ^* may be *random*. This is known in some application of bootstrap, regarding time series with unit roots (Basawa et al., 1991), inference at the boundary of parameter space (Andrews, 2000) and infinite variance processes (Cavaliere et al., 2016). In this context, randomness of τ^* gives an evidence of bootstrap failure, even in the first order. From a different point of view, this means that bootstrap p–values may be not asymptotically uniformly distributed as required in conventional bootstrap inference. Our objective is to connect bootstrap in instrumental variable estimation to this growing literature. In order to show the effects of randomness in bootstrap applications, we also consider the theoretical case of irrelevant instruments in TSLS (and IV) estimation, i.e. when $\boldsymbol{\pi} = \mathbf{0}$ (or $\pi = 0$), and associated t/Wald statistic under the null hypothesis $H_0 : \beta = \beta_0$. We also decide to analize separately IV and TSLS estimator to enlighten differences in the behaviour of bootstrapped distribution of estimators (and statistics), occurring when overidentication arises and the amount of (possibly unuseful) information increases. In this context, some empirical examples and simulation exercises, regarding the behaviour of bootstrap under irrelevant instruments, are proposed. We find that non–normality of estimators and test statistics also depends on the bootstrap method.

At the end of this Chapter, a small–scale Monte Carlo simulation is conducted in order to show the differences between bootstrap distribution of considered estimators and its conventional asymptotic version under weak instruments, and to understand issues regarding inference through bootstrap methods under poorly relevant instruments. Sample moments of (standardized) bootstrapped estimators are computed in order to highlight how non–normality emerges among a severe deterioration of instruments strength consider different levels of endogeneity. Furthermore we analyze performance of resampling methods in bias adjustment, hypothesis testing and confidence intervals to see in which cases inference could be misleads under low relevance, severe endogeneity and weakly endogenous instruments.

We introduce some notations: the symbols $\sim^*$ denotes the distribution in the bootstrap world, where "$\xrightarrow{d^*}_p$" denotes, weak convergence in probability. The symbol $\xrightarrow{d^*}_d$, indicates weak convergence in distribution conditionally on the data D_T (see Cavaliere and Georgiev, 2018). Finally, P^* is the probability measure induced by the bootstrap.

3.1 Bootstrap distribution under strong instruments

Under strong instruments asymptotics, i.e: $\boldsymbol{\pi} = \boldsymbol{\pi}_0 \neq \mathbf{0}$ where $\boldsymbol{\pi}_0$ is fixed, the bootstrap counterparts of the IV/TSLS estimators, denoted as $\hat{\beta}_T^{TSLS*}$ and $\hat{\beta}_T^{IV*}$, are (asymptotically) normally distributed, conditionally on the data $D_T = (\mathbf{y}, \mathbf{x}, \mathbf{Z})$. In the following paragraphs we present results regarding two types of residual–based resampling methods, where instruments may be resampled or not in the bootstrap DGP. This choice is related to the fact that both methods are used, in the IV literature, to obtain theoretical results under strong and weak instruments asymptotics, as mentioned in Section 2.4. It may be also useful to remark that both concentration parameter μ^2 and its bootstrap counterpart μ^{2*} diverge to ∞, under strong instruments asymptotics, as $T \to \infty$. Then, relevance of the instruments is preserved in the bootstrap world. We consider firstly the just identified case and then overidentified models.

3.1.1 IV estimator

Resampled Instrument

Considering $\pi = \pi_0 \neq 0$ where π_0 is fixed, the estimation error of bootstrap estimator defined in (2.45), using residual bootstrap with resampled instrument, is:

$$\hat{\beta}_T^{IV*} - \hat{\beta}_T^{IV} = (\mathbf{z}^{*\prime}\mathbf{x}^*)^{-1}\mathbf{z}^{*\prime}\mathbf{u}^* = \frac{\sum_{t=1}^T z_t^* u_t^*}{\sum_{t=1}^T z_t^* x_t^*}. \tag{3.1}$$

Furthermore, for $T \to \infty$ and $B \to \infty$:

$$\sqrt{T}(\hat{\beta}_T^{IV*} - \hat{\beta}_T^{IV}) \sim^* iid\left(0, \hat{\omega}^2\right),$$

where

$$\hat{\omega}^2 = \hat{\sigma}_u^2 \frac{(\sum_{t=1}^T x_t z_t)^2}{\sum_{t=1}^T z_t^2},$$

and could be also rewritten as:

$$\hat{\omega}^2 = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_z^2 \hat{\pi}_T^2}.$$

Therefore, $\hat{\sigma}_u^2 = T^{-1}\hat{\mathbf{u}}'\hat{\mathbf{u}}$, $\hat{\sigma}_z^2 = T^{-1}\sum_{t=1}^T (z_t - \overline{z})^2$ and finally $\hat{\omega}^2 - \omega^2 \xrightarrow{p} 0$. Hence, conditionally on the original data $D_T = (\mathbf{y}, \mathbf{x}, \mathbf{Z})$ the following asymptotic result holds:

$$\sqrt{T}(\hat{\beta}_T^{IV*} - \hat{\beta}_T^{IV}) \xrightarrow{d^*}_p N(0, \omega^2).$$

Fixed Instrument

Under strong instrument asymptotics, applying residual bootstrap with fixed instrument in the (bootstrap) DGP, the estimation error of bootstrap estimator defined in (2.42) is equal to:

$$\hat{\beta}_T^{IV*} - \hat{\beta}_T^{IV} = (\mathbf{z}'\mathbf{x}^*)^{-1}\mathbf{z}'\mathbf{u}^* = \frac{\sum_{t=1}^T z_t^* u_t^*}{\sum_{t=1}^T z_t^* x_t^*}. \tag{3.2}$$

Therefore, for $T \to \infty$ and $B \to \infty$:

$$\begin{aligned} \sqrt{T}(\hat{\beta}_T^{IV*} - \hat{\beta}_T^{IV}) &\sim N\left(0, \hat{\omega}^2\right) \\ &\xrightarrow{d^*}_p N\left(0, \hat{\omega}^2\right), \end{aligned} \tag{3.3}$$

where the variance $\hat{\omega}$ and other quantities are previously defined. Hence the expression in (3.2) converges, in a bootstrap sense, to a standard normal distribution:

$$\sqrt{T}\hat{\omega}^{-1}(\hat{\beta}_T^{IV*} - \hat{\beta}_T^{IV}) \xrightarrow{d^*}_p N(0,1).$$

3.1.2 Two Stage Least Squares

Resampled Instruments

When the number of instruments is more than one ($k > 1$) and under strong instrument asymptotics, i.e. $\boldsymbol{\pi} = \boldsymbol{\pi}_0 \neq \mathbf{0}$ where $\boldsymbol{\pi}_0$ is fixed, and applying residual bootstrap with resampled instruments, TSLS estimator, defined in (2.44), is:

$$\begin{aligned} \hat{\beta}_T^{TSLS*} &= \left(\mathbf{x}^{*\prime} P_{Z^*}\mathbf{x}^*\right)^{-1}\left(\mathbf{x}^{*\prime} P_{Z^*}\mathbf{y}^*\right) \\ &= \left[\mathbf{x}^{*\prime}\mathbf{Z}^*(\mathbf{Z}^{*\prime}\mathbf{Z}^*)^{-1}Z^{*\prime}\mathbf{x}^*\right]^{-1}\left[\mathbf{x}^{*\prime}\mathbf{Z}^*(\mathbf{Z}^{*\prime}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*\prime}\mathbf{y}^*\right]. \end{aligned}$$

Recalling that it could be rewritten in the following way:

$$\hat{\beta}_T^{TSLS*} = \hat{\beta}_T^{TSLS} + \left(\mathbf{x}^{*\prime} P_{Z^*}\mathbf{x}^*\right)^{-1}\left(\mathbf{x}^{*\prime} P_{Z^*}\mathbf{u}^*\right),$$

and its associated estimation error is $\hat{\beta}_T^{TSLS*} - \hat{\beta}_T^{TSLS}$, the bootstrap asymptotic distribution is the following:

$$\begin{aligned} \hat{\beta}_T^{TSLS*} - \hat{\beta}_T^{TSLS} &= \left(\mathbf{x}^{*\prime} P_{Z^*}\mathbf{x}^*\right)^{-1}\left(\mathbf{x}^{*\prime} P_{\mathbf{Z}^*}\mathbf{u}^*\right) \\ &= \left[\mathbf{x}^{*\prime}\mathbf{Z}^*(\mathbf{Z}^{*\prime}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*\prime}\mathbf{x}^*\right]^{-1}\left[\mathbf{x}^{*\prime}\mathbf{Z}^*(\mathbf{Z}^{*\prime}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*\prime}\mathbf{u}^*\right] \\ \sqrt{T}(\hat{\beta}_T^{TSLS*} - \hat{\beta}_T^{TSLS}) &\sim^* N(0, \hat{\omega}^2) \\ &\xrightarrow{d^*}_p N(0, \omega^2), \end{aligned}$$

where $\hat{\omega}^2 = \hat{\sigma}_u^2\left(\hat{\pi}_T' \frac{\mathbf{Z}'\mathbf{Z}}{T} \hat{\pi}_T\right)^{-1}$, $\omega^2 = \sigma_u^2\left[\boldsymbol{\pi_0}'\mathbf{Q}_{ZZ}\boldsymbol{\pi_0}\right]^{-1}$ and also $\hat{\omega}^2 - \omega^2 \to_p 0$, as in the just identified case.

Fixed Instruments

Under $\boldsymbol{\pi} = \boldsymbol{\pi}_0 \neq \mathbf{0}$ where $\boldsymbol{\pi}_0$ is fixed and applying residual–based bootstrap with fixed instruments ($k > 1$), the estimation error of bootstrap estimator, defined in (2.41), is:

$$\begin{aligned}\hat{\beta}_T^{TSLS*} - \hat{\beta}_T^{TSLS} &= \left(\mathbf{x}^{*\prime} P_Z \mathbf{x}^*\right)^{-1}\left(\mathbf{x}^{*\prime} P_Z \mathbf{u}^*\right) \\ &= \left[\mathbf{x}^{*\prime}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{x}^*\right]^{-1}\left[\mathbf{x}^{*\prime}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{u}^*\right].\end{aligned}$$

Furthermore, for $T \to \infty$ and $B \to \infty$:

$$\sqrt{T}(\hat{\beta}_T^{TSLS*} - \hat{\beta}_T^{TSLS}) \sim^* N(0, \hat{\omega}^2) \quad \text{where} \quad \hat{\omega}^2 = \hat{\sigma}_u^2\left[\hat{\pi}_T' \frac{\mathbf{Z}'\mathbf{Z}}{T}\hat{\pi}_T\right]^{-1}.$$

Hence, conditionally on the original data D_T:

$$\sqrt{T}(\hat{\beta}_T^{TSLS*} - \hat{\beta}_T^{TSLS}) \xrightarrow{d^*}_p N(0, \sigma_u^2\left[\boldsymbol{\pi_0}'\mathbf{Q}_{ZZ}\boldsymbol{\pi_0}\right]^{-1}).$$

To summarize, under strong instrument asymptotics, we have the following result for both considered resampling methods of IV/TSLS estimator (with resampled and fixed instruments):

$$\sqrt{T}\frac{(\hat{\beta}_T^* - \hat{\beta}_T)}{\hat{\omega}} \xrightarrow{d^*}_p N(0,1).$$

This means that conventional bootstrap asymptotics holds under the assumptions of strong instruments.

3.2 Bootstrap under weak instrument asymptotics: a new derivation

In this section we introduce a new derivation for the asymptotic distribution of $\hat{\beta}_T^{IV*}$ and $\hat{\beta}_T^{TSLS*}$ under weak instrument asymptotics, i.e. $\boldsymbol{\pi_0} = C/\sqrt{T}$ or $\pi = cT^{-1/2}$, applying residual–based bootstrap (with fixed and resampled instruments) under assumptions of homoskedastic disturbances and completely exogenous instruments. The main idea is to clarify the reason of bootstrap failures in IV that remains still under debate, as also pointed out in recent work, e.g. Andrews et al. (2018) and Young (2017).

As shown in Section 2.2, under a single irrelevant instrument the distribution of

IV estimator converges to a ratio of (correlated) normal variables. Moreover, the distribution of bootstrap IV counterpart converges, in a bootstrap sense, to a random distribution that differs from this derived Stock and Yogo (2005) by a quantity here denoted with the symbol $\mathbf{U}$. Using our set up, assumptions in Section 2.1.2 and imposing that estimated $\hat{\Sigma}$ converges in probability to a non–random matrix Σ, this quantity is also normally distributed but *conditionally* on the original data D_T. In overidentified cases the interpretation could be slightly less intuitive; random component appears both in numerator and denominator of the bootstrapped limiting distribution (under weak instruments), even conditionally to the data.

As mentioned in Section 2.2, under weak instrument asymptotics and normal disturbances (and instruments), population constant C (or c in the just identified case) plays a center role to guarantee also consistency and asymptotic normality of IV/T-SLS estimators, representing directly the strength of instruments. Its relevance is also confirmed in the bootstrap world; loosely speaking, when C (c) is high, the random components tend to disappear, vanishing when $C \to \infty$ reducing to strong instrument asymptotics. On the contrary, in the worst identification case of $c = 0$ (single irrelevant instrument), the randomness becomes predominant in the asymptotic distribution of bootstrapped IV estimator. These features (i.e. non–normality and randomness) may be found also in bootstrapped distribution of test statistics which are sensitive to weak instruments, as t/Wald or the Sargan statistic, computed using bootstrap estimator $\hat{\beta}_T^*$ or its associated resiudals $\hat{u}_t^* = y_t^* - \hat{\beta}_T^* x_t^*$. We point out that these results may be obtained applying other resampling methods, as for example parametric bootstrap introduced in Section 2.4.1 under strictly assumptions (normally and identical distributed disturbances), or the fully non–parametric pair bootstrap, very popular in IV setting. However, there are some exceptions.

A slightly computational analysis of the performance regarding bootstrap version of t–statistic under irrelevant instruments is also presented. Small–scale Monte Carlo exercises confirm that bootstrap counterpart of other limited information estimators included in $\kappa-$class, as LIML, present similar behaviour under poor relevance, although they are considered (partially) robust to weak instruments. Moreover, we do not provide a formal proof of the limiting distribution of other $\kappa-$class estimators.

3.2.1 Bootstrap IV under weak instrument asymptotics

Resampled Instrument

Theorem 1 (Bootstrapped IV estimator under weak instrument applying residual bootstrap with resampled instrument)**.** *Let the assumptions in Section 2.1.2 hold with $\pi = c/T^{1/2}$. Then, the distribution of bootstrapped IV estimator based on residual bootstrap with resampled instrument, conditionally on the data D_T, satisfies the following convergence as $T \to \infty$:*

$$\hat{\beta}_T^{IV*} - \hat{\beta}_T^{IV}|D_T \xrightarrow{d^*}_d \frac{w_{zu}^*}{c\sigma_z^2 + w_{zv}^* + \mathbf{U}}\bigg| \mathbf{U},$$

where $\mathbf{U} \sim N(0, \sigma_z^2\sigma_v^2)$, the symbol "$\cdot|D_T$" means conditionally on the sample D_T and $(w_{zu}^*, w_{zv}^*)' \sim N_2(0, \sigma_z^2\mathbf{\Sigma})$, independent of the original data, with $\mathbf{\Sigma}$ defined as in (2.6), and also $\hat{\mathbf{\Sigma}} \xrightarrow{p} \mathbf{\Sigma}$.

Fixed Instrument

Theorem 2 (Bootstrapped IV estimator under weak instrument applying residual bootstrap with fixed instrument)**.** *Let the assumptions in Section 2.1.2 hold with $\pi = c/T^{1/2}$. Then, the distribution of bootstrapped IV estimator based on residual bootstrap with fixed instrument z_t, conditionally on the data D_T, satisfies the following convergence as $T \to \infty$:*

$$(\hat{\beta}_T^{IV*} - \hat{\beta}_T^{IV})|D_T \quad \xrightarrow{d^*}_d \quad \frac{w_{zu}^*}{c\sigma_z^2 + w_{zv}^* + \mathbf{U}}\bigg|\mathbf{U}, \tag{3.4}$$

where the quantities $\mathbf{U}$, w_{zu}^* and w_{zv}^* are previously defined. Furthermore, in case of $\sigma_z^2 = 1$, the expressions in (1) and (3.4) is reduced to:

$$(\hat{\beta}_T^{*IV} - \hat{\beta}_T^*)|D_T \xrightarrow{d^*}_p \frac{w_{zu}^*}{c + w_{zv}^* + \mathbf{U}}\bigg|\mathbf{U},$$

where $\mathbf{U} \sim N(0, \sigma_v^2)$ conditionally on D_T. To summarize, the inconsistency of IV estimator is reflected in the bootstrap world. In fact, $P^*(\hat{\beta}_T^{IV*} - \hat{\beta}_T^{IV} \leq x)$ weakly converges to a probability law F depending on the random component $\mathbf{U}$, as follows:

$$P^*(\hat{\beta}^{IV*} - \hat{\beta}^{IV} \leq x) \xrightarrow{d} F(x, \mathbf{U})|\mathbf{U}.$$

3.2.2 Bootstrap TSLS under weak instrument asymptotics

Resampled Instruments

Theorem 3 (Bootstrapped TSLS estimator under weak instrument asymptotic applying residual bootstrap with resampled instruments)**.** *Let the assumptions in Section 2.1.2 hold with $\boldsymbol{\pi} = C/T^{1/2}$. Then, the distribution of bootstrapped TSLS estimator based on residual bootstrap with resampled instrument, conditionally on the data D_T, satisfies the following converges as $T \to \infty$:*

$$(\hat{\beta}_T^{TSLS*} - \hat{\beta}_T^{TSLS})|D_T \xrightarrow{d^*}_d \frac{(\mathbf{Q}_{ZZ}C + \mathbf{U} + W^*_{Zv})'\mathbf{Q}^{-1}_{ZZ}W^*_{Zu}}{(\mathbf{Q}_{ZZ}C + \mathbf{U} + W^*_{Zv})'\mathbf{Q}^{-1}_{ZZ}(\mathbf{Q}_{ZZ}C + \mathbf{U} + W^*_{Zv})}\Big|\mathbf{U}, \quad (3.5)$$

where $\mathbf{U} \sim N(0, \sigma_v^2\mathbf{Q}_{ZZ})$, conditionally on the data D_T, $\mathbf{Q}_{ZZ}$ is a full–rank squared matrix, $(W^*_{Zu}, W^*_{Zv})' \sim N(0, \mathbf{\Sigma} \otimes \mathbf{Q}_{ZZ})$ and again $\hat{\mathbf{\Sigma}} \xrightarrow{p} \mathbf{\Sigma}$.

Fixed Instruments

Theorem 4 (Bootstrapped TSLS estimator under weak instrument applying residual bootstrap with fixed instruments)**.** *Under the assumption of Section 2.1.2 and applying residual bootstrap with fixed instruments ($k > 1$), the distribution of bootstrap TSLS estimator, considering weak instrument asymptotics $\boldsymbol{\pi} = CT^{-1/2}$ and conditionally on the data D_T, converges to:*

$$(\hat{\beta}_T^{TSLS*} - \hat{\beta}_T^{TSLS})|D_T \xrightarrow{d^*}_d \frac{(\mathbf{Q}_{ZZ}C + \mathbf{U} + W^*_{Zv})'\mathbf{Q}^{-1}_{ZZ}W^*_{Zu}}{(\mathbf{Q}_{ZZ}C + \mathbf{U} + W^*_{Zv})'\mathbf{Q}^{-1}_{ZZ}(\mathbf{Q}_{ZZ}C + \mathbf{U} + W^*_{Zv})}\Big|\mathbf{U}, \quad (3.6)$$

where $\mathbf{U}$ and other $W^*_{Z\cdot}$ quantities are previously defined. Following expression (3.5) and (3.6), we observe that random component appears both in the numerator and denominator of TSLS limit bootstrap distribution. Proofs of Theorems 1,2,3, 4 are presented in Section 3.5.

3.2.3 Irrelevant instruments and bootstrap

In the no identification case, i.e. $\boldsymbol{\pi} = 0$ (or $\pi = 0$ when $k = 1$), β is not point identified and the distribution of IV/TSLS estimators results asymptotically as a ratio of two correlated normals (Cauchy–like random variable), as shown in (2.18) and (2.19). In this situation the bootstrap, exactly like asymptotic theory, completely breaks down[1]. This occurs because $\hat{\pi}_T$, obtained through OLS, estimates the true value ($\pi = 0$ in the just identified case) with probability equal to zero; loosely speaking, the unidentification is not properly captured in the bootstrap DGP. Nevertheless, since irrelevance of instruments represent only a theoretical limit case, it could be useful to understand the source of randomness in the limiting bootstrapped distribution of estimator, conditionally on the data. We further introduce the following corollary regarding bootstrapped IV estimator in the limit case of $\pi = c = 0$.

Corollary 4.1 (Bootstrapped IV estimator under irrelevant instrument applying residual bootstrap)**.** *When $c = \pi = 0$, given the assumption of Section 2.1.2, the bootstrap estimator $\hat{\beta}_T^{IV*}$ based on residual bootstrap (using fixed or resampled instrument), conditionally on the sample D_T, converges to:*

- *Resampled z_t*

$$(\hat{\beta}_T^{IV*} - \hat{\beta}_T^{IV})|D_T \overset{d^*}{\to}_d \frac{w_{zu}^*}{w_{zv}^* + \mathbf{U}}\Big|\mathbf{U}. \tag{3.7}$$

- *Fixed z_t*

$$\begin{aligned}(\hat{\beta}_T^{IV*} - \hat{\beta}_T^{IV})|D_T &= (\mathbf{z}'\mathbf{x}^*)^{-1}\mathbf{z}'\hat{\mathbf{u}}^* \\ &\overset{d^*}{\to}_d \frac{w_{zu}^*}{w_{zv}^* + \mathbf{U}}\Big|\mathbf{U}. \end{aligned}\tag{3.8}$$

The distribution in (3.7) and (3.8) is also a Cauchy–like random variable, but differs from the asymptotic IV under irrelevant instrument. In fact, recalling expression in (2.18), when $\sigma_u^2 = \sigma_v^2 = 1$, the distribution of IV under unidentification is equal to w_{zu}/w_{zv}. This distribution is centered on the OLS plim, i.e. the endogeneity coefficient ρ, as pointed out in simulation study of Section 2.3.

Furthermore, a straightforward Monte Carlo exercise is conducted in order to understand and visualize randomness of $\hat{\beta}^{IV*}$ under two residual–based resampling methods. Figure 3.1 shows 10 empirical densities[2], obtained through $B = 9999$ residual–based bootstrap samples with (left panel) and without (right panel) resampled instrument. In the simulation design, z_t is drawn from a standard normal, the stochastic

[1] Applying estimators and test non–robust to weak instruments, like IV, TSLS and its associated t–stats

[2] Densities are computed applying normal Kernel on the bootstrap replications $\tilde{\beta}_{Tb}^* \in (-4, 4)$, for $b = 1, \ldots, B$. The bandwidth h is selected using Silverman's Rule of Thumb (1986), and it is approximately equal to $h \approx 1.06\hat{sd}(\hat{\beta}_T^{IV})T^{-1/5}$

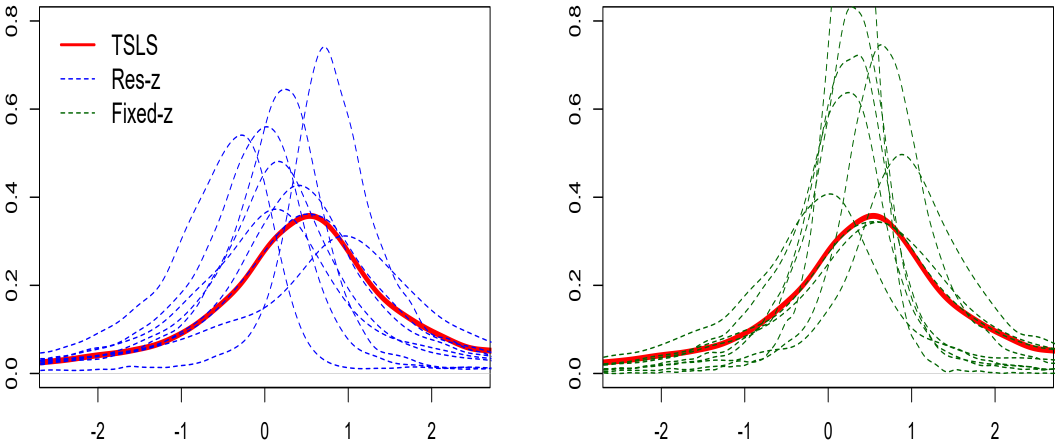
components are jointly normally distributed and true value of parameter is set to $\beta = 0$. The sample size is moderately large ($T = 250$), where the considered level of endogeneity is high and equal to $\rho = Cor(u_t, v_t) = 0.9$. Red line represents the asymptotic distribution (Cauchy–like) of IV estimators, based on $M = 100000$ simulated samples, while blue (resampled instruments), and green (fixed) dashed lines are the bootstrapped IV estimator when $c = \mu = 0$. Despite IV estimator under $\pi = 0$ is centered on ρ, its bootstrapped counterpart substantially differs from its finite sample distribution, even if two resampling methods seem to perform quite different. In fact, bootstrap distributions with resampled instrument are centered on random values, often very far from 0.5, while fixed–instrument distributions seem to present an high kurtosis.

In overidentified models there is an increase of useless information; thus, randomness appears less severe, especially using residual bootstrap with resampled instruments. In Figure 3.2, we plot only 10 bootstrap–based densities of TSLS with $k = 5$ (multi–normal instruments) and high degree of endogeneity level ($\rho = 0.9$), where $\boldsymbol{\pi} = \mathbf{0}$. Even if the (simulated) asymptotic distribution of TSLS under five irrelevant instruments is again centered on $\rho = 0.9$, the bootstrap distributions of estimators, generated with the same DGP present very different medians from this theoretical value.

3.2.4 Concentration parameter and bootstrap under weak instruments

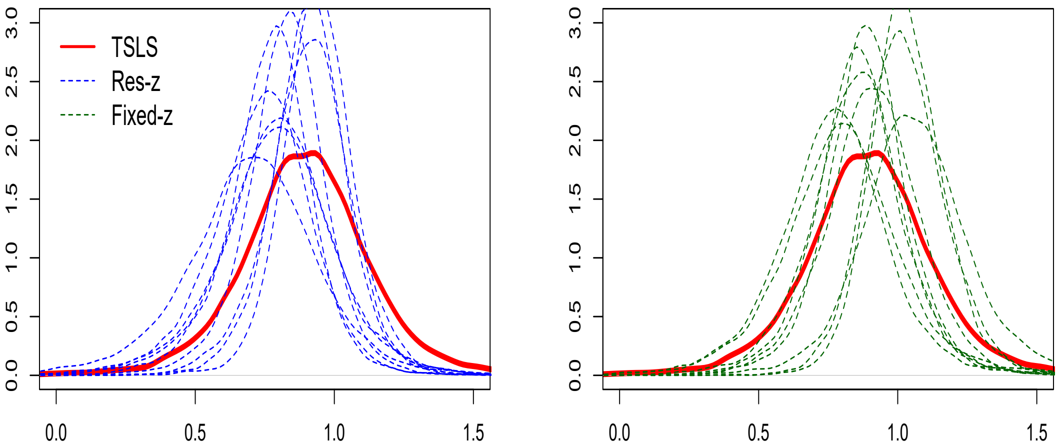
In this section we report some results from Zhan (2017), regarding the difference between population concentration parameter and its bootstrapped counterpart converging to different constant as $T, B \to \infty$ under weak instrument asymptotics. Here, for simplicity, we present only the just identified case. When $\pi = c/\sqrt{T}$, the bootstrap counterpart of (population) concentration parameter, denoted as μ^{*2}, converges to following distribution as $T \to \infty$:

$$\begin{aligned}
\mu^{*2} &= \frac{\hat{\sigma}_z^2 \hat{\pi}_T^2}{\hat{\sigma}_v^2} \\
&\xrightarrow{d} \frac{c^2 \sigma_z^2 + 2c w_{zu} + (w_{zv})^2/\sigma_z^2}{\sigma_v^2},
\end{aligned}$$



(a) Resampled z_t (b) Fixed z_t

Figure 3.1: Empirical density of 10 bootstrapped IV estimators under irrelevant instrument ($\pi = 0$) using $B = 9999$ replications.



(a) Resampled Instruments (b) Fixed Instruments

Figure 3.2: Empirical densitiy of 10 bootstrap–based TSLS (residual bootstrap) estimators under $k = 5$ irrelevant instruments ($\boldsymbol{\pi} = 0$) using $B = 9999$ replications.

while[3] $\mu^2 \xrightarrow{p} \sigma_z^2 c^2/\sigma_v^2$. Hence the asymptotic difference between μ^{2*} and μ^2 is equal to:

$$\frac{2cw_{zu} + (w_{zv})^2}{\sigma_z^2}.$$

For this reason, we are interested in comparing the estimated strength of instruments in the bootstrap world against its asymptotic counterpart, in order to check or quantify this difference. We generate $M = 1000$ samples of x_t, z_t under weak instrument asymptotics following the design of Section 2.3.2. Then we apply residual–based bootstrap proposed in this chapter (both fixed and resampled instruments). We consider only just identified case, reporting Monte Carlo estimate of μ^2, computed as $\hat{\mu}^2 = M^{-1}\sum_{m=1}^{M} F_m - 1$, where F is obtained through (2.28), and the rejection frequencies of $F > 10$ test for different strength of identification, i.e. μ^2/k varies from 0 to 60. The bootstrapped estimators or the concentration parameter are computed in the following way:

$$\hat{\mu}^{*2} = \hat{E}(F^*) - 1 = B^{-1}\sum_{b=1}^{B} F_b^* - 1,$$

where F_b^* obtained under resampled and fixed instruments are:

$$F_b^* = \frac{\hat{\sigma}_z^{*2}\pi_T^{*2}}{\hat{\sigma}_v^2} \; F^* = \frac{\hat{\sigma}_z^2\pi_T^{*2}}{\hat{\sigma}_v^2}.$$

We compute mean and median values of $\hat{\mu}^{*2}$ and the median of $F_b^* > 10$ for $b = 1, \ldots, B$. In Table 3.3, we observe that the mean and median of bootstrap estimator of μ^2 are greater than μ^2 and $\hat{\mu}^2$ in the irrelevant instrument case. In general, under weak instrument asymptotics, the bootstrap overestimates in mean the concentration parameter, while the distance between $\hat{\mu}^2$ and μ^2 decreases in large sample sizes ($T = 500$). To summarize, the difference between concentration parameter and its bootstrapped counterpart appears not negligible only if the instrument is irrelevant or very weak, i.e. the DGP presents $\mu^2 \leq 1$. Further investigations may be conducted using different DGPs (e.g. non–normal disturbances and invalid instrument).

[3] Remark the Appendix of Section 2

3.2.5 Cautionary note: a (parametric) fixed regressors bootstrap

In the previous paragraphs, two residual–based resampling scheme are applied in order to prove convergence (in a bootstrap sense) of bootstrapped IV/TSLS estimators to a non–normal and random distribution, exploiting a particular sequence of modeling (weak instrument asymptotics). From a practical perspective, non–normality of bootstrap estimators could be a desirable feature in order to check the presence of weak instruments, allowing to develop useful indicators and forma diagnostics bootstrap–based tests. For example, it would be possible to test normality of bootstrap replications, or associated statistics (under the null or the alternative hypothesis). However, considering another resampling method the distribution of a bootstrapped statistic may be asymptotically normal even if instruments are collectively weak (i.e. μ^2/k or $\pi \approx 0$) or totally irrelevant. An example is represented by the parametric fixed design i.i.d. bootstrap in the context of IV estimator (perfectly identified case). Applying this method, the bootstrap DGP may be summarized as follows:

$$y_t^\star = \hat{\beta}_T^{IV} x_t + u_t^\star \text{ where } u_t^\star \sim N(0, \hat{\sigma}_u^2). \tag{3.9}$$

Following the expression in (3.9), both endogenous regressor and instrument are fixed at their sample original values. Then, bootstrapped IV estimator takes the following form:

$$\hat{\beta}_T^{IV\star} = \frac{\sum_{t=1}^T z_t y_t^\star}{\sum_{t=1}^T z_t x_t} = \hat{\beta}_T^{IV} + \frac{\sum_{t=1}^T z_t u_t^\star}{\sum_{t=1}^T z_t x_t}, \tag{3.10}$$

and the associated estimation error in the bootstrap world is:

$$\begin{aligned} \hat{\beta}_T^{IV\star} - \hat{\beta}_T^{IV} &= (\mathbf{z}'\mathbf{x})^{-1}\mathbf{z}\mathbf{u}^\star \\ &= \frac{\sum_{t=1}^T z_t u_t^\star}{\sum_{t=1}^T z_t x_t}. \end{aligned}$$

Hence, conditionally on the original data D_T, following theorem proves asymptotic normality of bootstrap estimator, obtained through fixed regressor parametric bootstrap, under weak instruments.

Theorem 5 (Bootstrapped IV estimator under weak instruments applying fixed iid bootstrap)**.** *Under weak instrument asymptotics $\pi = T^{-1/2}c$ where $0 \leq c < \infty$, assumption in Section 2.1.2 and applying method described in* (3.9) *and estimator in*

(3.10), *the bootstrapped distribution of IV, conditionally on the data D_T, converges to:*

$$\sqrt{T}\hat{\omega}^{-1}(\hat{\beta}_T^{*IV} - \hat{\beta}_T^{IV}) \xrightarrow{d^*}_p N(0,1). \tag{3.11}$$

Proof is presented in Section 3.5, i.e. the Appendix of this Chapter. To summarize, method described in (3.9) neglets $\hat{\pi}_T$ and residuals v_t in the bootstrap DGP, while in general these quantities bring out identification issues from the sample to bootstrap world through the inconsistent estimator $\hat{\beta}_T^{IV}$.

Table 3.1 presents main results of a small–scale Monte Carlo exercise; in this context we apply Jarque–Bera normality tests directly on the distribution of $\hat{\beta}_T^{IV\star}$, i.e. $(\hat{\beta}_{T1}^{IV\star}, \ldots, \hat{\beta}_{TB}^{IV\star})$ obtained through (parametric) fixed regressor iid bootstrap. The number of generated samples is $M = 1000$; in each iteration the number of bootstrap replication is equal to $B = 199$. We notice that rejection frequencies of JB tests are very low and close to nominal level $\alpha = 0.05$ for all degrees of identification, indicated by μ^2, and two considered levels of endogeneity. We hightlight in bold highly p–values and denote in red the best "performance" in terms of size. Empirical size is also found close to the nominal level using five instruments (TSLS) and similar rejections can be obtained with higher values of B; all of these results are consistent with Theorem 5.

Moreover, applying the same bootstrap method, other $\kappa-$class estimators could present a different behaviour with respect to IV/TSLS, becoming non–normal under weak or irrelevant instruments. This can be viewed in Table 3.2, presenting rejection frequency of JB test applied on the bootstrapped distribution of the LIML estimator, even with $M = 1000$ and $B = 199$. In the unidentified case ($\boldsymbol{\pi} = \mathbf{0}$), rejection frequencies are close to one with $k = 2, 5$ instruments[4], whereas they tend to the nominal level if the strength of identification increases. This empirical evidence appears under the two considered level of endogeneity. To summarize, this may depend on two things. On one hand the discrepancy between $\hat{\kappa}_{LIML}$ and $\hat{\kappa}_{LIML}^*$ could be reflected in the bootstrap counterpart, while in TSLS estimation $\hat{\kappa} = \hat{\kappa}^* = 1$. On the other hand, recalling the definition in Section 2.15, LIML takes always account of quantities in the first stage.

Thus, the combination between (bootstrap) method and estimator is required in this setting: a particular method may be chosen in order to check weak identification only if it effectively preserves patterns of low relevance (i.e. non–normality) in asymptotic distribution of estimator, conditionally on the data D_T. Finally, we point out

[4]We recall that if $k = 1$ LIML coincides with TSLS

$k = 1$

ρ	$\mu^2 = 0$	$\mu^2 = 1$	$\mu^2 = 10$	$\mu^2 = 20$
0.5	0.044	0.050	0.035	0.041
0.9	0.048	0.040	0.048	0.032

$k = 5$

ρ	$\mu^2/k = 0$	$\mu^2/k = 1$	$\mu^2/k = 10$	$\mu^2/k = 20$
0.5	0.044	0.038	0.046	0.042
0.9	0.040	0.040	0.046	0.044

Table 3.1: Rejection frequency of JB test applied the bootstrapped IV/TSLS ($k = 1, 5$) distribution, based on fixed regressor iid bootstrap, considering four different levels of identification

$k = 2$

ρ	$\mu^2/k = 0$	$\mu^2/k = 1$	$\mu^2/k = 10$	$\mu^2/k = 20$
0.5	**0.968**	0.872	0.126	0.044
0.9	0.956	0.832	0.124	0.050

$k = 5$

ρ	$\mu^2/k = 0$	$\mu^2/k = 1$	$\mu^2/k = 10$	$\mu^2/k = 20$
0.5	0.980	0.824	0.038	0.040
0.9	**0.988**	0.812	0.058	0.034

Table 3.2: Rejection frequency of JB test on the bootstrapped LIML distribution computed with iid fixed regressor bootstrap for different levels of identification

that fixed regressor bootstrap is not valid under weak (or irrelevant) instruments and in general its application seems not appropriate in IV estimation.

3.2.6 Bootstrapped t–statistic under irrelevant instrument

Randomness and non–normality of bootstrap limiting distribution may appear even if inference is conducted through t–test for the null-hypotesis $H_0 : \beta = \beta_0$. To show this issue, we recall the simplest IV case, imposing $\sigma_u = \sigma_v = 1$ and $\sigma_{uv} = \rho$. Under irrelevant instrument i.e. $\pi = 0$, the t–statistic converges in distribution to the following expression:

$$\tau_T \xrightarrow{d} \frac{w_{zu}/w_{zv}}{\sqrt{1 - 2\rho\frac{w_{zu}}{w_{zv}} + \left(\frac{w_{zu}}{w_{zv}}\right)^2}}. \tag{3.12}$$

Expression in (3.12), comes from the fact that estimator $\hat{\sigma}_u^2$ could be no longer consistent under irrelevant instruments (see for example the textbook of B. Hansen, 2018), converging in distribution to a non–trivial limit, equal to $\Psi = 1 - 2\rho(w_{zu}/w_{zv}) + (w_{zu}/w_{zv})^2$ and violating assumptions of Section 2.1.2. Hence, under an irrelevant instrument, the distribution of t–statistic under the null hypothesis is non–normal and the parameter ρ plays a key role, as we point out in simulation study of Section 2.3. Considering for simplicity only residual bootstrap with fixed instruments, following Davidson and MacKinnon (2010), the bootstrap counterpart of t–statistics is equal to: $\tau_T^* = T^{1/2}\hat{\omega}^{*-1}(\hat{\beta}_T^* - \hat{\beta}_T)$ previously defined in expression (2.46). In this case $\hat{\beta}_T$ could be IV/TSLS (or other $\kappa-$class estimator) and $T^{1/2}\hat{\omega}^*$ denotes the standard error of β_T^* in the bootstrap world, obtained with the same method of t–statistic which is bootstrapped. In just identified case, the simplest choice is $\hat{\omega}^{*2} = T^{-1/2}\sqrt{\hat{\sigma}_u^{*2}\sum_{t=1}^T z_t^2}/\sum_{t=1}^T z_t x_t^*$, but also robust (to heteroskedasticity) estimators[5] may be used.

In Figure 3.3, we plot 10 empirical distribution of bootstrapped t–statistic with $B = 9999$, applying the simulation design of Section 3.2.5 applying residual bootstrap with fixed and resampled instruments. In this exercise, the true value of parameter β is set equal to 0 and also the (high) endogeneity level is $\rho = 0.9$. The black dotted line represents the asymptotic distribution of the statistic under the null hypothesis $H_0 : \beta = 0$ (standard normal), while red line denotes the finite sample distribution of the t statistic, obtained as: $T^{1/2}\hat{\omega}^{-1}(\hat{\beta}_T^{IV} - \beta)$ where $\hat{\omega}/\sqrt{T}$ is the estimated standard error of $\hat{\beta}_T$. Thus, under the assumption of irrelevant (and weak) instruments, the asymptotic results $\tau_T^* \xrightarrow{d^*}_p N(0,1)$ no longer holds.

3.3 Bootstrap inference under weak instruments: a simulation study

In this section, we present the main result of Monte Carlo simulation conducted to evaluate non-normality and randomness, conditionally of the data, of bootstrapped dis-

[5] See for example White(1980) or Mackinnon and White (1985)

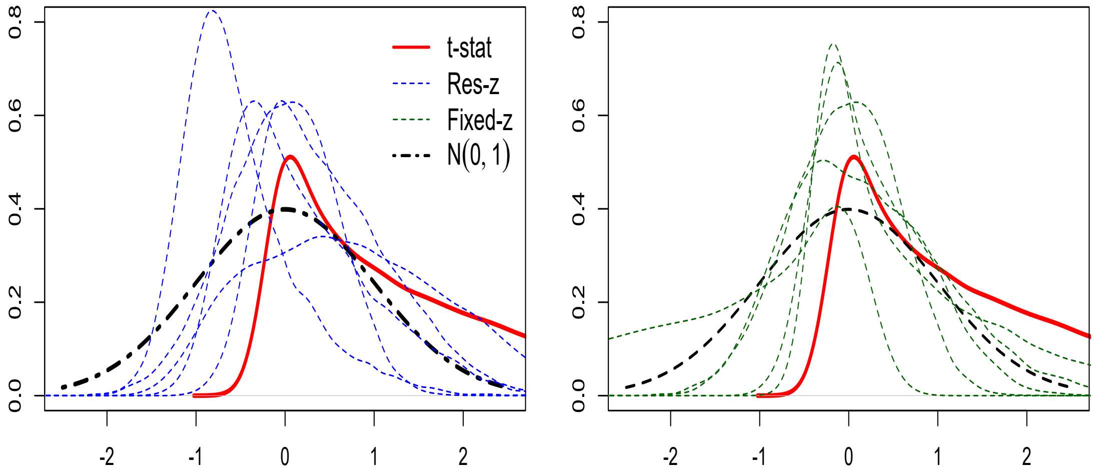


Figure 3.3: Empirical densities of bootstrap t-statistics (residual bootstrap) associated to IV under the null hypothesis $H_0 : \beta = 0$ and irrelevant instrument, i.e. $\pi = 0$.

tribution of IV/TSLS (and other $\kappa-$class estimators) under weak instruments. From a different point of view, our concern is also to understand if sources of non-normality could help to detect weak instruments. For this purpose, the presence of random component, denoted with the symbol $\mathbf{U}$, could be a serious issue, especially under nearly irrelevant instruments, as highlighted in simulation exercise of Section 3.2.3.

First of all, different bootstrap methods surveyed in Chapter 2 are proposed to find if they present substantial differences against asymptotic distribution of estimator. The analysis includes $\kappa-$class estimators and weak exogeneity of instruments. Secondly, we present performance of bootstrap methods regarding several areas: bias correction, hypothesis testing and confidence intervals. The design for Monte Carlo simulation basically relies on that proposed in Section 2.3.2, considering different strength of instruments imposed by the average concentration parameter μ^2/k and several levels of endogeneity for the regressor. Sample size varies from 100 to 1000, while the number of bootstrap replication is $B = 199, 399$.

3.3.1 Non normality of bootstrap distribution

As mentioned in Section 3.2, the bootstrapped distribution of IV/TSLS estimator may be random and highly non–normal under weak instrument asymptotics. In order to show these features, we are interested in quantifying how the standardized bootstrap counterpart of estimator is far from the standard normal, computing sample mean, median, variance, skewness, kurtosis and InterQuartile Range (IQR) of $(\tilde{\beta}^*_{T1}, \ldots, \tilde{\beta}^*_{TB})$, where $\tilde{\beta}^* = \sqrt{T}\hat{\omega}^{-1}(\hat{\beta}^* - \hat{\beta})$. The number of bootstrap replications is $B = 399$ for each of $M = 1000$ random samples. We present both average and median values across the M replicated samples, denoting all values with the symbol '*'. The main idea is to investigate if malfunctions of conventional standardization can be sistematically found in the bootstrap samples of estimators, due to a) the inconsistency of $\hat{\beta}_T$ and b) the subsequentely randomness of $\hat{\beta}^*_T$. Again, we use bold and red color to emphasize worst and best performances.

Table 3.4 contains some results regarding the case of $k = 3$ instruments: average values of mean and median of $\left\{\tilde{\beta}^*_{Tb}\right\}_{b=1}^{B}$ increase with the endogeneity level while the average kurtosis of bootstrapped distribution suggests that non-normality may occur especially when μ^2/k is less than 10. Average and median of Kurtosis (of bootstrapped estimator) could take huge values in presence of very weak instruments, when $\mu^2/k = 1$, due to the no–moment problem. If the sample size increases, as shown in Table 3.5, standardized TSLS estimator performs better, on average, in terms of IQR, variance and kurtosis of the bootstrapped distribution. In this setting the number of instruments is $k = 5$.

κ-class estimator

Simulation includes some results regarding bootstrapped $\kappa-$class estimators, obtained applying residual–based bootstrap method. The resampling scheme is the same used for TSLS, involving residuals induced by LIML or Fuller($\underline{c}$) estimates i.e. $\hat{u}_t = y_t - \hat{\beta}^{LIML}_T x_t$ and $\hat{u}_t = y_t - \hat{\beta}^{Fuller}_T x_t$, where $\hat{\beta}^{LIML}_T$ and $\hat{\beta}^{Fuller}_T$ are introduced in Section 2.1.4 through expression (2.15). Simulation design follows that proposed in this Section.

Table 3.6 shows the main results in a overidentified case ($k = 5$ instruments), considering Fuller estimator with constant $\underline{c}$ equal to 1. Standardized bootstrapped LIML presents unreliable huge variance and kurtosis under weak instruments, performing very well in terms of Median when instruments are collectively not weak, i.e. $\mu^2/k = 10$. For this reason, a huge variability in bootstrap distribution of the LIML could be interpreted as a signal of weak identification although its sensitivity, producing those outliers, depends on the so called no–moment problem. Fuller estimator seems to outperform LIML and TSLS especially in terms of skewness and kurtosis.

3.3.2 Bootstrap–based bias correction

In order to show mulfunctions of bootstrap inference under weak instrument asymptotics, we present results regarding bias–correction of the TSLS estimators[4]. The bootstrap–based bias corrected version of TSLS is computed as follows:

$$\hat{\beta}_T^{BTSLS} = 2\hat{\beta}_T^{TSLS} - \frac{1}{B}\sum_{b=1}^{B}\hat{\beta}_{Tb}^{TSLS*} \tag{3.13}$$

where the quantity $B^{-1}\sum_{b=1}^{B}\hat{\beta}_{Tb}^{TSLS*}$ represents the bootstrap estimates of the bias $E(\hat{\beta}_T^{TSLS}) - \beta$ of the estimator, obtained as the average of the B bootstrap replications. Evaluation is conducted through descriptive indicators previously introduced (RMSE, MAE, IDR, KS) also using graphical inspection against the normal distribution. In addition, coverage rates are obtained through (bootstrapped) gaussian confidence intervals:

$$\hat{\beta}^{TSLS} \pm \Phi(1-\alpha/2)\hat{\sigma}^*,$$

where the symbol $\hat{\sigma}^{*2}$, denotes the bootstrap estimates of the variance (of the estimator) as follows:

$$\hat{\sigma}^{*2} = \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\beta}_{T,b}^{TSLS*} - B^{-1}\sum_{b=1}^{B}\hat{\beta}_{T,b}^{TSLS*}\right)^2.$$

In Figure 3.4 we plot empirical density of bootstrap–based bias corrected TSLS under four different strength of identification and four degrees of endogeneity, against the normal density (red line) with mean equal to $\beta = 0$ and variance equal to $V(\hat{\beta}) = T^{-1}\sigma_u^2(\boldsymbol{\pi}'\mathbf{Q}_{ZZ}\boldsymbol{\pi})^{-1} = T^{-1}(\boldsymbol{\pi}'I_k\boldsymbol{\pi})^{-1}$, considering a number of instruments equal to $k = 3, 5$. When $k = 3$, and $\mu^2/k = 1$, the distribution of bias corrected TSLS could be far from $N(\beta, V(\hat{\beta}_T))$ even in the case of low endogeneity ($\rho = 0.25$), indicated with the black line. Therefore, in case of strong instruments ($\mu^2/k = 40$), the distribution

[4]Perfectly identified case (IV estimator) is not considered because bias of estimators is mainly affected by overidentification in finite samples

seems to coincide with its reference normal under all levels of endogeneity. Figure 3.5 confirms some issue under very weak instruments for each level of ρ.

Table 3.7 shows the performance of bootstrap–based bias corrected TSLS when $k = 5$. In terms of confidence intervals, covera1ge rates reach the nominal value only in case of low endogeneity, resulting understated in other cases. Kolmogorov Smirnov statistic KS, computed on $M = 1000$ replicated samples, takes values less than 0.05 even if the instruments present low μ^2/k. Table 3.8 presents the performance of bootstrap–based bias correted LIML, where the estimator $\hat{\beta}_T^{BLIML}$ using $\hat{\beta}_T^{LIML}$ and its bootstrap counterpart $\hat{\beta}_T^{LIML*}$. It seems to perform better than bootstrap–based bias corrected TSLS especially in terms of coverage rates when instruments are not too weak ($\mu^2/k \geq 10$), resulting usual unreliable huge values of RMSE under nearly irrelevant and weak instruments, due to its lack of all moments.

3.3.3 Confidence intervals

In this subsection we discuss confidence intervals of IV/TSLS under weak instruments. As previously illustrated in Section 2.3, coverage rates may be severely affected by low relevance of instruments and this reflects in the bootstrap world. We consider two types of non–robust (with respect to weak instruments) bootstrap–based sets. The first is the so called *percentile* and is computed as:

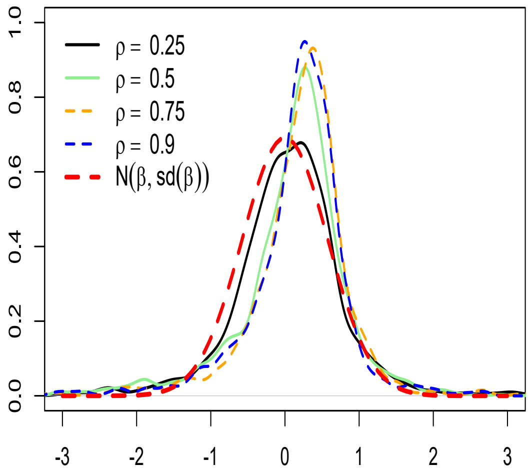
$$CI^*_{P,1-\alpha} = (\hat{\beta}^*_{T,\alpha/2}, \hat{\beta}^*_{T,1-\alpha/2}), \tag{3.14}$$

where $\hat{\beta}^*_T$ is bootstrap counterpart of IV/TSLS (or $\kappa-$class) estimator, and $\hat{\beta}^*_{T,j}$ represent the j–percentile obtained through B replication. The latter confidence interval is expressed as follows:

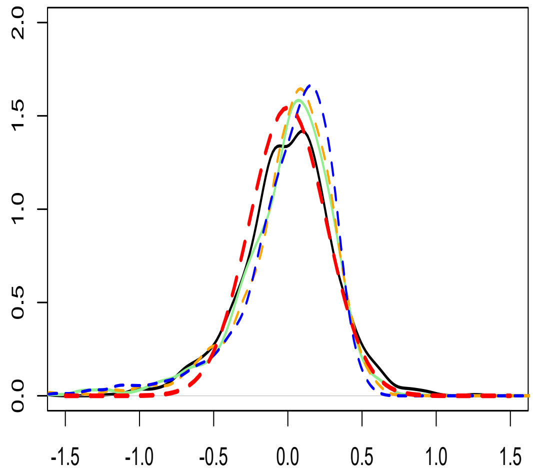
$$CI^*_{t,1-\alpha} = (\hat{\beta}_T - \tau^*_{\alpha/2} \cdot \hat{\omega}/\sqrt{T}, \hat{\beta}_T - \tau^*_{1-\alpha/2} \cdot \hat{\omega}/\sqrt{T}), \tag{3.15}$$

were $\tilde{\omega}$ is the estimated standard error of $\hat{\beta}_T$, $t^*_\alpha, t^*_{1-\alpha}$ are the estimated $\alpha/2$ and $1-\alpha/2$ quantiles of the distribution of τ^*_T. The sets in (3.15) are denoted as t–bootstrap confidence intervals (or t–percentile), suggested by Davidson and MacKinnon (2014). We point out that other confidence sets can be obtained inverting a fully robust test but they are not considered here.

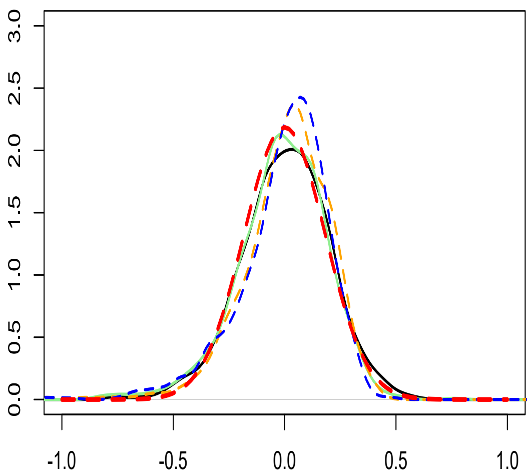
Table 3.9 contains the median length and coverage rates of $M = 1000$ confidence



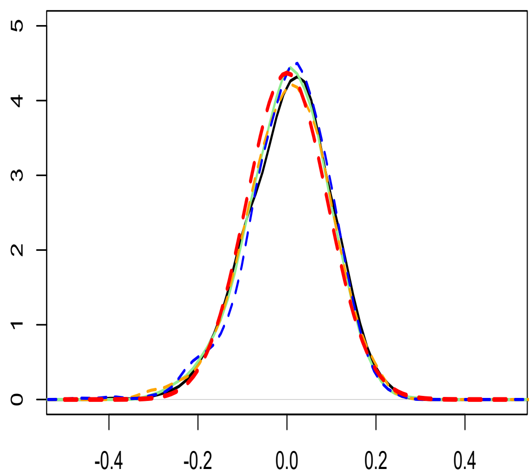
(a) $\mu^2/k = 1$



(b) $\mu^2/k = 5$

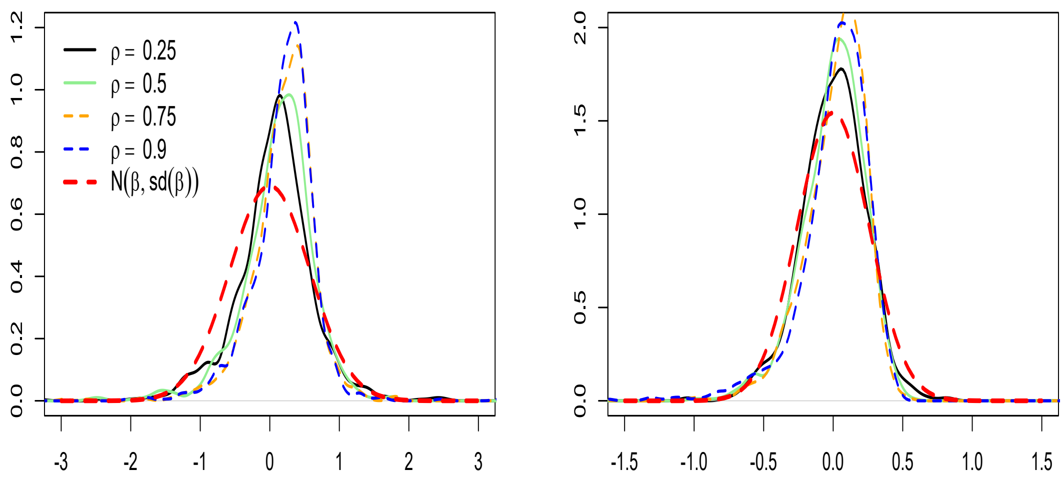


(c) $\mu^2/k = 10$

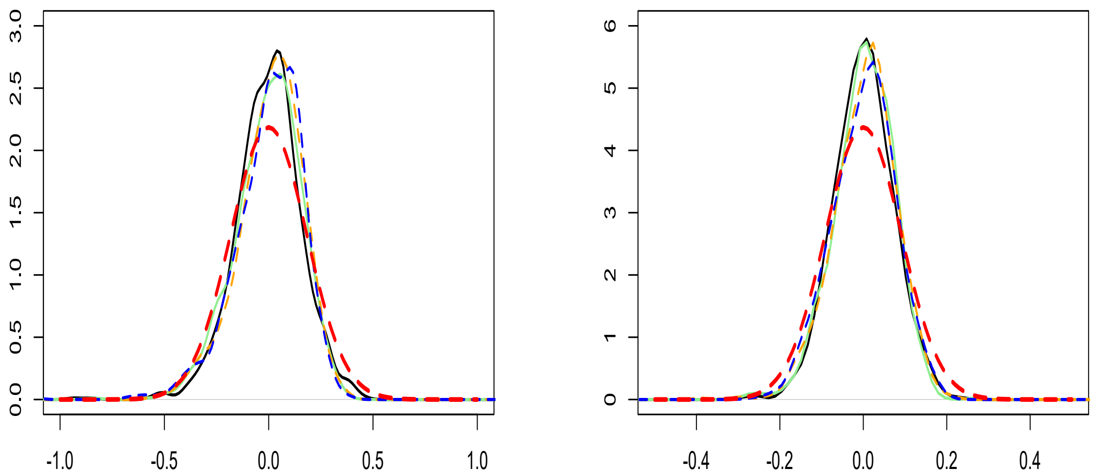


(d) $\mu^2/k = 40$

Figure 3.4: Empirical density of bootstrap–based bias corrected TSLS estimator ($k = 3$) under different degrees of identification and levels of endogeneity.



(a) $\mu^2/k = 1$ (b) $\mu^2/k = 5$



(c) $\mu^2/k = 10$ (d) $\mu^2/k = 40$

Figure 3.5: Empirical density of bootstrap–based bias corrected TSLS estimator ($k = 5$) with different degrees of identification and levels of endogeneity.

sets obtained using asymptotic theory and method described in (3.14) and (3.15), with a number of bootstrap replication equal to $B = 399$, sample size $T = 100$ and also $\beta = 0$. We consider only two levels of endogeneity, i.e. $\rho = 0.5, 0.9$. Under weak instruments, when $k = 1$, percentile intervals are wider than those obtained with conventional inference, presenting coverage rate close to one under unidentification. Sets obtained through t–boot method perform very poor in terms of coverage rates, especially if the instruments are not too weak (i.e. $\mu^2/k = 10$), although they have lower length when instruments are very weak or irrelevant with respet to the asymptotic sets. This is related to the randomness of bootstrapped t–statistic, preivously viewed in Section 3.2.6. Nevertheless, in the overidentified case ($k = 5$) both two bootstrap methods have very poor performance in terms of coverage rates even when instruments are not too weak, i.e. $\mu^2/k \leq 20$ and under high endogeneity, while t–boot sets perform better than percentiles under strong instruments.

Furthermore we show results regarding (non robust) confidence sets LIML estimator applying both Pair and Residual bootstrap. Main results are in Table 3.10. Under no–identification all confidence intervals are very large (in median), confirming no–moment problem occurring in LIML case. Sets obtained through residual method are slightly narrower than those computed with the pairs one. However, coverage rates of percentile intervals are very different between two methods and Pair seems to perform better when instruments are not very strong ($\mu^2/k = 10$). Surprisingly, where identification level is very high, bootstrap confidence intervals may perform worst than those obtained though asymptotic methods, especially under high endogeneity.

3.3.4 Wald test under weak instruments

In this subsection we analize the performance of bootstrapped t–test for the null hypothesis of $H_0 : \beta = \beta_0$, where in our analysis we consider $\beta_0 = 0$. We use two unrestricted methods (Pair and Unrestricted Residual) that do not impose the null hypothesis in the bootstrap DGP, as viewed in Section 2.4. When instruments are irrelevant the rejection frequencies (empirical size) of bootstrap–based test increase with the number of instruments, as shown in Figure 3.6. Therefore, if overidentification is combined with high endogeneity, rejection frequencies tend rapidely to unit even if the true DGP involves $\beta_0 = 0$. When instruments are collectively not weak, bootstrap methods could be helpful to improve the performance, especially in presence of high endogeneity, as highlighted in the lower panel of Figure 3.6. Figure 3.7 and 3.8 compare asymptotic p–values of the test with those obtained through residual bootstrap. We consider two different scenarios: $\mu^2/k = 1$, i.e. weak instruments, and $\mu^2/k = 20$, i.e. strong relevance, running $M = 500$ simulations with $\rho = 0.5$. In case of a single weak instrument (red points), bootstrap p–values are sistematically lower than those obtained through conventional asymptotics; in overidentified cases this discrepancy still exists but is less systematic.

Table 3.11 contains rejection frequency of t–statistics using two bootstrap methods and asymptotics considering a number of instruments equal to $k = 1, 5$. The number of bootstrap replication is $B = 399$ (Monte Carlo are 1000) and $\beta = 0$. In the just identified case, when a single instrument is weak or irrelevant, the bootstrap performs better than asymptotics; under moderate endogeneity residual bootstrap performs better than pair in terms of empiriacal size, expect when endogeneity is too high. Moreover, in overidentified case (i.e. $k = 5$), Pair seems to outperform residual bootstrap. Table 3.12 presents similar results regarding empirial size of bootstrapped t–test under a *weakly* endogenous instrument, following the design of Section 2.3. We notice that Pair bootstrap performs better than Residual when the level of endogeneity is not too high ($\phi = 0.1$). A possible reason is that bootstrap DGP (residual case) is constructed under the strictly hypothesis of incorrelation between instruments and structural disturbances. When endogeneity of instrument increases (i.e. $\phi = 1$) all methods get worse for all considered scenarios of identification and endogeneity (of the regressor); they are very far from the nominal level, confirming problems discussed in simulation of Chapter 2. A possible challenge in this topic is to apply restricted methods proposed by Davidson and MacKinnon (2010) illustrated in Section 2.4. We write an R code presented in the Appendix useful to implement these two methods.

3.4 Concluding Remarks

In this Chapter we analyze the properties of bootstrapped distribution of estimators IV models with one endougenous regressor. We propose a new derivation of bootstrapped IV/TSLS estimator under weak instruments asymptotics, using two types of residual–based bootstrap, involving resampled or fixed instruments. We show that this distribution has a random limit, conditionally on the data, and also confirms asymptotical non–normality; these features vanish as the correlation between instruments and the endogenous regressor increases. In particular, bootstrapped distribution of

IV/TSLS differs from those obtained under weak instrument asymptotics, even if it preserves some peculiarities of low relevance. Randomness clearly appears in the limit case of irrelevant instruments, considering both estimator and associated t–statistic. This framework could be extended to the case of multiple endogenous regressors, even if the vector of nuisance parameter $\mathbf{\Pi}$ becomes a $k \times m$ matrix, where m is the number of rhs endogenous variables and the strength of instruments is represented by concentration matrix $\boldsymbol{\mu}^{\mathbf{2}}/T$.

The method can be also straightforward applied to prove the limiting distribution of bootstrapped $\kappa-$class estimators under poorly relevant instruments. It is well known that (partially) robust estimators, as LIML and Fuller, are asymptotically non–normal under weak instruments (as previously illustrated in Chapter 2). Therefore, some bootstrap methods are deemed invalid in this context. Empirical exercise suggests that bootstrap $\kappa-$class estimators may also present a random limiting distribution, conditionally on the data, when instruments are irrelevant or very weak. However, a possible drawback is that parameter κ could be a random variable in such cases, e.g. LIML and Fuller, rather than being a constant in OLS and IV/TSLS.

Inference conducted via bootstrap methods could be dramatically affected by weak instruments, as viewed through Monte Carlo simulation. In particular, bootstrap–based bias corrected TSLS estimator may be non–normal, in finite samples, under weak instruments and high endogeneity. Confidence sets can present wrong coverage rates when weak identification is combined with lack of moments, while rejection frequencies of bootstrapped t–Wald test are severely affected by overidentification and non exogeneity of instruments. In this context, we show that usage of t–boot confidence intervals may not help to improve performance in terms of coverage rates and rejection frequency under very weak instruments; moreover, residuals based bootstrap may performs worst than Pair bootstrap if high endogeneity is combined with overidentification. In this context, will apply the so called fast double bootstrap to see if this method can improve performance of t/Wald tests under misspecification.

3.5 Appendix

Bootstrap distribution of IV under weak instruments

The bootstrap estimator of IV (residual bootstrap with resampled instrument) is defined as follows:

$$\begin{aligned}\hat{\beta}_T^{IV*} - \hat{\beta}_T^{IV} &= (\mathbf{z}^{*\prime}\mathbf{x}^*)^{-1}(\mathbf{z}^{*\prime}\mathbf{u}^*) \\ &= \frac{\sum_{t=1}^T z_t^* u_t^*}{\sum_{t=1}^T z_t^* x_t^*}\end{aligned}$$

Under the assumption of Section 2.1.2 and conditionally on the original data, the bootstrap instruments and disturbances, $(z_t^*, u_t^*, v_t^*)'$, are i.i.d. with mean zero and covariance matrix

$$\begin{bmatrix}\hat{\sigma}_z^2 & 0 & 0 \\ 0 & \hat{\sigma}_u^2 & \hat{\sigma}_{uv} \\ 0 & \hat{\sigma}_{uv} & \hat{\sigma}_v^2\end{bmatrix}$$

where, by assumption,

$$\begin{bmatrix}\hat{\sigma}_z^2 & 0 & 0 \\ 0 & \hat{\sigma}_u^2 & \tilde{\sigma}_{uv} \\ 0 & \hat{\sigma}_{uv} & \hat{\sigma}_v^2\end{bmatrix} \xrightarrow{p} \begin{bmatrix}\sigma_z^2 & 0 & 0 \\ 0 & \sigma_u^2 & \sigma_{uv} \\ 0 & \sigma_{uv} & \sigma_v^2\end{bmatrix}$$

Moreover, since the term $z_t^* u_t^*$ is (conditionally on the original data) i.i.d. with mean zero, provided that u_t and z_t have finite fourth order moments, we can proceed as e.g. in Liu et al. (1988) and prove the (conditional) central limit theorem:

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^T z_t^* u_t^* &\sim {}^*\text{iid}(0, \hat{\sigma}_z^2\hat{\sigma}_u^2) \xrightarrow{d^*}_p N(0, \sigma_u^2\sigma_z^2) \\ \frac{1}{\sqrt{T}}\sum_{t=1}^T z_t^* v_t^* &\sim {}^*\text{iid}(0, \hat{\sigma}_z^2\hat{\sigma}_v^2) \xrightarrow{d^*}_p N(0, \sigma_v^2\sigma_z^2).\end{aligned}$$

where "$\xrightarrow{d^*}_p$" denotes weak convergence in probability. The above convergence results are joint; that is,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} z_t^* \begin{pmatrix} u_t^* \\ v_t^* \end{pmatrix} \xrightarrow{d^*}_p N\left(0, \sigma_z^2 \mathbf{\Sigma}\right) = \begin{pmatrix} w_{zu}^* \\ w_{zv}^* \end{pmatrix}$$

where

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix}$$

and w_{zu}^* and w_{zv}^* are implicitly defined. Consider the denominator, and recall that under weak instrument asymptotics, i.e. $\pi = c/\sqrt{T}$, the OLS estimator of π, given by

$$\hat{\pi}_T = \frac{\sum_{t=1}^{T} z_t x_t}{\sum_{t=1}^{T} z_t^2} = \frac{c}{\sqrt{T}} + \frac{\sum_{t=1}^{T} z_t v_t}{\sum_{t=1}^{T} z_t^2},$$

satisfies

$$\begin{aligned} \sqrt{T}(\hat{\pi}_T - \frac{c}{\sqrt{T}}) &= \sqrt{T}\hat{\pi}_T - c = \frac{T^{-1/2}\sum_{t=1}^{T} z_t v_t}{T^{-1}\sum_{t=1}^{T} z_t^2} \\ &\xrightarrow{d} N(c, \sigma_v^2/\sigma_z^2). \end{aligned}$$

The denominator bootstrap IV estimator can be rewritten as:

$$\frac{1}{T}\sum_{t=1}^{T} z_t^* x_t^* = \hat{\pi}_T \frac{1}{T}\sum_{t=1}^{T} z_t^{*2} + \frac{1}{T}\sum_{t=1}^{T} z_t^* v_t^*,$$

where

$$\frac{1}{T}\sum_{t=1}^{T} z_t^{*2} \xrightarrow{p^*}_p \sigma_z^2, \tag{3.16}$$

that is, for any $\delta > 0$, $P^*(|\frac{1}{T}\sum_{t=1}^{T} z_t^{*2} - \sigma_z^2| > \delta) \xrightarrow{p} 0$ (P^* is the probability measure induced by the bootstrap). Moreover, since $z_t^* v_t^*$ is conditionally i.i.d., $\frac{1}{T}\sum_{t=1}^{T} z_t^* v_t^* = O_p^*\left(T^{1/2}\right)$ in probability (see Chang and Park, 2003, for the definitions of bootstrap stochastic orders). All together, as in Cavaliere and Georgiev (2018, proof of Theorem 3), these results imply that, conditionally on the data,

$$(\hat{\beta}_T^{IV*} - \hat{\beta}_T^{IV})|D_T \xrightarrow{d^*}_d \frac{w_{zu}^*}{c\sigma_z^2 + \mathbf{U} + w_{zv}^*}\Big|\mathbf{U}.$$

as required.

Hence, conditionally on the fact that $\mathbf{U} \sim N(0, \sigma_z^2 \sigma_v^2)$, the bootstrapped asymptotic distribution of IV estimator could be rewritten as follows:

$$\begin{aligned} \frac{w_{zu}}{c\sigma_z^2 + \mathbf{U} + w_{zv}} &= \frac{\sigma_u^2 \sigma_z^2 N(0,1)}{c\sigma_z^2 + \sigma_z^2 \sigma_v^2 N(0,1) + \sigma_z^2 \sigma_v^2 N(0,1)} \\ &= \frac{\sigma_u^2 N(0,1)}{c + \sigma_v^2 N(0,1) + \sigma_v^2 N(0,1)} \\ &= \frac{\sigma_u^2 N(0,1)}{c + \sigma_v^2 N(0,2)}, \end{aligned}$$

where two normals are correlated with correlation coefficient equal to ρ.

Bootstrapped IV estimator with fixed instrument

The bootstrap counterpart of IV estimator applying residual bootstrap with fixed instruments is:

$$\begin{aligned} \hat{\beta}_T^{IV*} - \hat{\beta}_T &= (\mathbf{z}'\mathbf{x}^*)^{-1}(\mathbf{z}'\mathbf{u}^*) \\ &= \frac{\sum_{t=1}^T z_t u_t^*}{\sum_{t=1}^T z_t x_t^*} \end{aligned}$$

Again, under the assumption of Section 2.1.2 and conditionally on the original data, the bootstrap disturbances, $(u_t^*, v_t^*)'$, are i.i.d. with mean zero and covariance matrix

$$\hat{\mathbf{\Sigma}} = \begin{bmatrix} \hat{\sigma}_u^2 & \hat{\sigma}_{uv} \\ \hat{\sigma}_{uv} & \hat{\sigma}_v^2 \end{bmatrix}$$

where, by assumption,

$$\begin{bmatrix} \hat{\sigma}_u^2 & \hat{\sigma}_{uv} \\ \hat{\sigma}_{uv} & \hat{\sigma}_v^2 \end{bmatrix} \rightarrow_p \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix}$$

Here, $z_t u_t^*$ is also (conditionally on the original data) i.i.d. with mean zero, provided that:

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{t=1}^T z_t u_t^* &\sim {}^*\text{iid}(0, \sigma_z^2 \hat{\sigma}_u^2) \xrightarrow{d^*}_p N(0, \sigma_u^2 \sigma_z^2) \\ \frac{1}{\sqrt{T}} \sum_{t=1}^T z_t v_t^* &\sim {}^*\text{iid}(0, \sigma_z^2 \hat{\sigma}_v^2) \xrightarrow{d^*}_p N(0, \sigma_v^2 \sigma_z^2). \end{aligned}$$

The above convergence results are the following:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} z_t \begin{pmatrix} u_t^* \\ v_t^* \end{pmatrix} \xrightarrow{d^*}_p N\left(0, \sigma_z^2 \mathbf{\Sigma}\right) = \begin{pmatrix} w_{zu}^* \\ w_{zv}^* \end{pmatrix}$$

where

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix}$$

and w_{zu}^* and w_{zv}^* are previously defined. Under weak instrument asymptotics, i.e. $\pi = c/\sqrt{T}$, the OLS estimator of π, given by

$$\hat{\pi}_T = \frac{\sum_{t=1}^{T} z_t x_t}{\sum_{t=1}^{T} z_t^2} = \frac{c}{\sqrt{T}} + \frac{\sum_{t=1}^{T} z_t v_t}{\sum_{t=1}^{T} z_t^2},$$

satisfies

$$\begin{aligned} \sqrt{T}(\hat{\pi}_T - \frac{c}{\sqrt{T}}) &= \sqrt{T}\hat{\pi}_T - c = \frac{T^{-1/2}\sum_{t=1}^{T} z_t v_t}{T^{-1}\sum_{t=1}^{T} z_t^2} \\ &\xrightarrow{d} N(c, \sigma_v^2/\sigma_z^2). \end{aligned}$$

The denominator of bootstrap IV estimator could be written

$$\frac{1}{T}\sum_{t=1}^{T} z_t x_t^* = \hat{\pi}_T \frac{1}{T}\sum_{t=1}^{T} z_t^2 + \frac{1}{T}\sum_{t=1}^{T} z_t v_t^*,$$

where $T^{-1}\sum_{t=1}^{T} z_t^2 \xrightarrow{p} \sigma_z^2$ by assumptions. Since $z_t v_t^*$ is conditionally i.i.d., $\frac{1}{T}\sum_{t=1}^{T} z_t v_t^* = O_p^*\left(T^{1/2}\right)$ in probability these results imply that, conditionally on the data,

$$(\hat{\beta}_T^{IV*} - \hat{\beta}_T^{IV})|D_T \xrightarrow{d^*}_d \frac{w_{zu}^*}{c\sigma_z^2 + \mathbf{U} + w_{zv}^*}\Big|\mathbf{U},$$

as required.

Considering both fixed and resampled instrument, in the special unidentified case of irrelevant instrument $\pi = c = 0$:

$$(\hat{\beta}_T^{IV*} - \hat{\beta}_T^{IV})|D_T \xrightarrow{d^*}_d \frac{w_{zu}}{w_{zv} + \mathbf{U}}\Big|\mathbf{U}, \tag{3.17}$$

and then expression in (3.17) could be straightforward rewritten as:

$$\begin{aligned}
\frac{w_{zu}}{w_{zv} + \mathbf{U}} &= \frac{N(0, \sigma_z^2 \sigma_u^2)}{N(0, \sigma_z^2 \sigma_v^2) + N(0, \sigma_z^2 \sigma_v^2)} \\
&= \frac{N(0, \sigma_z^2 \sigma_u^2)}{N(0, 2\sigma_z^2 \sigma_v^2)},
\end{aligned}$$

resulting, even conditionally on original data D_T, a ratio of two correlated normal random variables.

More than one instrument

Resampled Instruments

We firstly report the expression for bootstrap counterpart of TSLS estimator under residual–based bootstrap with resampled instruments (overidentified case, $k > 1$):

$$\begin{aligned}
\hat{\beta}_T^{*TSLS} &= \left(\mathbf{x}^{*\prime} P_{Z^*} \mathbf{x}^*\right)^{-1} \left(\mathbf{x}^{*\prime} P_{Z^*} \mathbf{y}^*\right) \\
&= \left[\mathbf{x}^{*\prime} \mathbf{Z}^* (\mathbf{Z}^{*\prime} \mathbf{Z}^*)^{-1} \mathbf{Z}^{*\prime} \mathbf{x}^*\right]^{-1} \left[\mathbf{x}^{*\prime} \mathbf{Z}^* (\mathbf{Z}^{*\prime} \mathbf{Z}^*)^{-1} \mathbf{Z}^{*\prime} \mathbf{y}^*\right] \\
&= \hat{\beta}_T^{TSLS} + \left(\mathbf{x}^{*\prime} P_{Z^*} \mathbf{x}^*\right)^{-1} \left(\mathbf{x}^{*\prime} P_{Z^*} \mathbf{u}^*\right) \qquad (3.18)
\end{aligned}$$

The joint distribution of bootstrapped residuals and instruments is:

$$\begin{pmatrix} Z_t^* \\ u_t^* \\ v_t^* \end{pmatrix} \sim^* \text{iid}\left(0, \mathbf{\Sigma}_T \otimes T^{-1}(\mathbf{Z}'\mathbf{Z})\right)$$

where $\hat{\mathbf{\Sigma}} = \begin{bmatrix} \hat{\sigma}_u^2 & \hat{\sigma}_{uv} \\ \hat{\sigma}_{uv} & \hat{\sigma}_v^2 \end{bmatrix}$. Then, we have the following asymptotic results:

$$\begin{aligned}
\frac{\mathbf{Z}^{*\prime}\mathbf{u}^*}{\sqrt{T}} &\sim {}^*\text{iid}_k(0, \hat{\sigma}_u^2 T^{-1}(\mathbf{Z}'\mathbf{Z})) \xrightarrow{d^*}_p N_k(0, \sigma_u^2 \mathbf{Q}_{ZZ}) \\
\frac{\mathbf{Z}^{*\prime}\mathbf{v}^*}{\sqrt{T}} &\sim {}^*\text{iid}_k(0, \hat{\sigma}_v^2 T^{-1}(\mathbf{Z}'\mathbf{Z})) \xrightarrow{d^*}_p N_k(0, \sigma_v^2 \mathbf{Q}_{ZZ})
\end{aligned}$$

Under weak instruments, i.e. $\boldsymbol{\pi} = CT^{-1/2}$ where C is a $k-$dimensional vector, the distribution of $\hat{\pi}_T$ is

$$\begin{aligned}
\hat{\pi}_T &= (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{x} \\
&= CT^{-1/2} + (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{v} \\
&\xrightarrow{d} N_k(C, \sigma_u^2 \mathbf{Q}_{ZZ}^{-1}).
\end{aligned} \tag{3.19}$$

Combined these results, the denominator of (3.18) becomes:

$$\begin{aligned}
\mathbf{x}^{*\prime} P_{Z^*}\mathbf{x}^* &= (\mathbf{x}^{*\prime}\mathbf{Z}^*)(\mathbf{Z}^{*\prime}\mathbf{Z}^*)^{-1}(\mathbf{Z}^{*\prime}\mathbf{x}^*) \\
&= (\mathbf{Z}^*\hat{\pi}_T + \mathbf{v}^*)'\mathbf{Z}^*(\mathbf{Z}^{*\prime}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*\prime}(\mathbf{Z}^*\hat{\pi}_T + \mathbf{v}^*) \\
&= \left(\frac{(\mathbf{Z}^*\hat{\pi}_T + \mathbf{v}^*)'\mathbf{Z}^*}{\sqrt{T}}\right)\left(\frac{\mathbf{Z}^{*\prime}\mathbf{Z}^*}{T}\right)^{-1}\left(\frac{\mathbf{Z}^{*\prime}(\mathbf{Z}^*\hat{\pi}_T + \mathbf{v}^*)}{\sqrt{T}}\right) \\
&= \left(\frac{\hat{\pi}_T\mathbf{Z}^{*\prime}\mathbf{Z}^*}{\sqrt{T}} + \frac{\mathbf{v}^{*\prime}\mathbf{Z}^*}{\sqrt{T}}\right)\left(\frac{\mathbf{Z}^{*\prime}\mathbf{Z}^*}{T}\right)^{-1}\left(\frac{\mathbf{Z}'\mathbf{Z}\hat{\pi}_T}{\sqrt{T}} + \frac{\mathbf{Z}^{*\prime}\mathbf{v}^*}{\sqrt{T}}\right) \\
&= \left(\frac{\sqrt{T}\hat{\pi}_T\mathbf{Z}^{*\prime}\mathbf{Z}^*}{\sqrt{T}\sqrt{T}} + \frac{\mathbf{v}^{*\prime}\mathbf{Z}^*}{\sqrt{T}}\right)\left(\frac{\mathbf{Z}^{*\prime}\mathbf{Z}^*}{T}\right)^{-1}\left(\frac{\mathbf{Z}'\mathbf{Z}\hat{\pi}_T\sqrt{T}}{\sqrt{T}\sqrt{T}} + \frac{\mathbf{Z}^{*\prime}\mathbf{v}^*}{\sqrt{T}}\right) \\
\mathbf{x}^{*\prime} P_{Z^*}\mathbf{x}^* | D_T &\overset{d^*}{\rightarrow}_d (\mathbf{Q}_{ZZ}C + \mathbf{U} + W^*_{Zu})'\mathbf{Q}_{ZZ}^{-1}(\mathbf{Q}_{ZZ}C + \mathbf{U} + W^*_{Zu})\Big|\mathbf{U},
\end{aligned} \tag{3.20}$$

where expression in (3.20) comes from:

$$\begin{aligned}
\mathbf{x}^{*\prime} P_{Z^*}\mathbf{x}^* &\overset{d^*}{\rightarrow}_d \left[N_k(C, \sigma_v^2\mathbf{Q}_{ZZ})Q_{ZZ} + W^*_{Zu}\right]'\mathbf{Q}_{ZZ}^{-1}\left[N_k(C, \sigma_v^2\mathbf{Q}_{ZZ})\mathbf{Q}_{ZZ} + W^*_{Zu}\right] \\
&= \left[C\mathbf{Q}_{ZZ} + \mathbf{U} + W^*_{Zu}\right]'\mathbf{Q}_{ZZ}^{-1}\left[CQ_{ZZ} + \mathbf{U} + W^*_{Zu}\right].
\end{aligned}$$

Furthermore, the numerator of expression (3.18) has the same asymptotic distribution, conditionally on the data:

$$
\begin{aligned}
\mathbf{x}^{*\prime} P_{Z^*}\mathbf{u}^* &= (\mathbf{x}^{*\prime}\mathbf{Z}^*)(\mathbf{Z}^{*\prime}\mathbf{Z}^*)^{-1}(\mathbf{Z}^{\prime}\mathbf{u}^*) \\
&= \frac{(\mathbf{x}^{*\prime}\mathbf{Z}^*)}{\sqrt{T}}\frac{(\mathbf{Z}^{*\prime}\mathbf{Z}^*)^{-1}}{T}\frac{(\mathbf{Z}^{*\prime}\mathbf{u}^*)}{\sqrt{T}} \\
&= \frac{(\mathbf{Z}^*\hat{\pi}_T+\mathbf{v})^{\prime}\mathbf{Z}}{\sqrt{T}}\frac{(\mathbf{Z}^{\prime}\mathbf{Z})^{-1}}{T}\frac{\mathbf{Z}^{\prime}\mathbf{u}^*}{\sqrt{T}} \\
&= \frac{(\mathbf{Z}\hat{\pi}_T+\mathbf{v}^*)^{\prime}\mathbf{Z}}{\sqrt{T}}\frac{(\mathbf{Z}^{\prime}\mathbf{Z})^{-1}}{T}\frac{\mathbf{Z}^{\prime}\mathbf{u}^*}{\sqrt{T}} \\
&= \left(\frac{\hat{\pi}_T\mathbf{Z}^{\prime}\mathbf{Z}}{\sqrt{T}}+\frac{\mathbf{v}^{*\prime}\mathbf{Z}}{\sqrt{T}}\right)\left(\frac{\mathbf{Z}^{\prime}\mathbf{Z}}{T}\right)^{-1}\left(\frac{\mathbf{Z}^{\prime}\mathbf{u}^*}{\sqrt{T}}\right) \\
&= \left(\frac{\sqrt{T}\hat{\pi}_T\mathbf{Z}^{\prime}\mathbf{Z}}{\sqrt{T}\sqrt{T}}+\frac{\mathbf{v}^{*\prime}\mathbf{Z}}{\sqrt{T}}\right)\left(\frac{\mathbf{Z}^{\prime}\mathbf{Z}}{T}\right)^{-1}\left(\frac{\mathbf{Z}^{\prime}\mathbf{u}^*}{\sqrt{T}}\right) \\
\mathbf{x}^{*\prime} P_{Z^*}\mathbf{u}^*|D_T &\overset{d^*}{\to}_d (C^{\prime}\mathbf{Q}_{ZZ}+\mathbf{U}+W^*_{Zu})^{\prime}\mathbf{Q}_{ZZ}^{-1}W^*_{Zu}\Big|\mathbf{U},
\end{aligned}
$$

where

$$
\mathbf{x}^{*\prime} P_{Z^*}\mathbf{u}^*|D_T \overset{d^*}{\to}_d \left[N_k(C,\sigma_v^2\mathbf{Q}_{ZZ})\mathbf{Q}_{ZZ}+W_{ZV}\right]^{\prime}\mathbf{Q}_{ZZ}^{-1}W^*_{Zu}\Big|\mathbf{U}
$$

and $(W^*_{Zu}, W^*_{Zu})^{\prime} \sim N(\mathbf{0},\Sigma\otimes\mathbf{Q}_{ZZ})$

Fixed instruments

The expression for bootstrap counterpart of TSLS estimator under residual–based bootstrap with fixed instruments ($k>1$) is:

$$
\begin{aligned}
\hat{\beta}^{*TSLS} &= (\mathbf{x}^{*\prime}P_Z\mathbf{x}^*)^{-1}(\mathbf{x}^{*\prime}P_Z\mathbf{y}^*) \\
&= \left[\mathbf{x}^{*\prime}\mathbf{Z}(\mathbf{Z}^{\prime}\mathbf{Z})^{-1}\mathbf{Z}^{\prime}\mathbf{x}^*\right]^{-1}\left[\mathbf{x}^{*\prime}\mathbf{Z}(\mathbf{Z}^{\prime}\mathbf{Z})^{-1}\mathbf{Z}^{\prime}\mathbf{y}^*\right] \\
&= \hat{\beta}_T^{TSLS}+(\mathbf{x}^{*\prime}P_Z\mathbf{x}^*)^{-1}(\mathbf{x}^{*\prime}P_Z\mathbf{u}^*)\,. \qquad (3.21)
\end{aligned}
$$

Bootstrap disturbances and fixed instruments have the following joint asymptotic distributions:

$$\begin{aligned}
\frac{\mathbf{Z}'\mathbf{u}^*}{\sqrt{T}} &\sim {}^*\mathrm{NID}(0, \hat{\sigma}_u^2 T^{-1}(\mathbf{Z}'\mathbf{Z})) \overset{d^*}{\rightarrow}_p N_k(0, \sigma_u^2 \mathbf{Q}_{ZZ}) \\
\frac{\mathbf{Z}'\mathbf{v}^*}{\sqrt{T}} &\sim {}^*\mathrm{NID}(0, \hat{\sigma}_v^2 T^{-1}(\mathbf{Z}'\mathbf{Z})) \overset{d^*}{\rightarrow}_p N_k(0, \sigma_v^2 \mathbf{Q}_{ZZ}).
\end{aligned}$$

Therefore, the denominator of expression (3.21) becomes:

$$\begin{aligned}
\mathbf{x}^{*\prime} P_Z \mathbf{x}^* &= (\mathbf{X}^{*\prime}\mathbf{Z})(\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{x}^*) \\
&= (\mathbf{Z}\hat{\pi}_T + \mathbf{v}^*)'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{Z}\hat{\pi}_T + \mathbf{v}^*) \\
&= \left(\frac{(\mathbf{Z}\hat{\pi}_T + \mathbf{v}^*)'\mathbf{Z}}{\sqrt{T}}\right)\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{-1}\left(\frac{\mathbf{Z}'(\mathbf{Z}\hat{\pi}_T + \mathbf{v}^*)}{\sqrt{T}}\right) \\
&= \left(\frac{\hat{\pi}_T\mathbf{Z}'\mathbf{Z}}{\sqrt{T}} + \frac{\mathbf{v}^{*\prime}\mathbf{Z}}{\sqrt{T}}\right)\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{-1}\left(\frac{\mathbf{Z}'\mathbf{Z}\hat{\pi}_T}{\sqrt{T}} + \frac{\mathbf{Z}'\mathbf{v}^*}{\sqrt{T}}\right) \\
&= \left(\frac{C/\sqrt{T}\mathbf{Z}'\mathbf{Z}}{\sqrt{T}} + \frac{\mathbf{Z}'\mathbf{v}}{\sqrt{T}} + \frac{\mathbf{v}^{*\prime}\mathbf{Z}}{\sqrt{T}}\right)\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{-1} \\
&\cdot \left(\frac{\mathbf{Z}'\mathbf{Z}C/\sqrt{T}}{\sqrt{T}} + \frac{\mathbf{Z}'\mathbf{v}}{\sqrt{T}} + \frac{\mathbf{Z}'\mathbf{v}^*}{\sqrt{T}}\right) \\
\mathbf{x}^{*\prime} P_Z \mathbf{x}^* | D_T &\overset{d^*}{\rightarrow}_d (\mathbf{Q}_{ZZ}C + \mathbf{U} + W_{Zu}^*)'\mathbf{Q}_{ZZ}^{-1}(\mathbf{Q}_{ZZ}C + \mathbf{U} + W_{Zu}^*)\Big|\mathbf{U} \qquad (3.22)
\end{aligned}$$

And the denominator of expression (3.21) becomes:

$$
\begin{aligned}
\mathbf{x}^{*\prime} P_Z \mathbf{u}^* &= (\mathbf{x}^{*\prime}\mathbf{Z})(\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{u}^*) \\
&= \frac{(\mathbf{x}^{*\prime}\mathbf{Z})}{\sqrt{T}} \frac{(\mathbf{Z}'\mathbf{Z})^{-1}}{T} \frac{(\mathbf{Z}'\mathbf{u}^*)}{\sqrt{T}} \\
&= \frac{(\mathbf{Z}\hat{\pi}_T + \mathbf{v}^*)'\mathbf{Z}}{\sqrt{T}} \frac{(\mathbf{Z}'\mathbf{Z})^{-1}}{T} \frac{\mathbf{Z}'\mathbf{u}^*}{\sqrt{T}} \\
&= \frac{(\mathbf{Z}\hat{\pi}_T + \mathbf{v}^*)'\mathbf{Z}}{\sqrt{T}} \frac{(\mathbf{Z}'\mathbf{Z})^{-1}}{T} \frac{\mathbf{Z}'\mathbf{u}^*}{\sqrt{T}} \\
&= \left(\frac{\hat{\pi}_T\mathbf{Z}'\mathbf{Z}}{\sqrt{T}} + \frac{\mathbf{v}^{*\prime}\mathbf{Z}}{\sqrt{T}}\right)\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{-1}\left(\frac{\mathbf{Z}'\mathbf{u}^*}{\sqrt{T}}\right) \\
&= \left(\frac{(C/\sqrt{T} + (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{v})\mathbf{Z}'\mathbf{Z}}{\sqrt{T}} + \frac{\mathbf{v}^{*\prime}\mathbf{Z}}{\sqrt{T}}\right)\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{-1}\left(\frac{\mathbf{Z}'\mathbf{u}^*}{\sqrt{T}}\right) \\
&= \left(\frac{(C/\sqrt{T}\mathbf{Z}' + \mathbf{Z}'\mathbf{v})}{\sqrt{T}} + \frac{\mathbf{v}^{*\prime}\mathbf{Z}}{\sqrt{T}}\right)\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{-1}\left(\frac{\mathbf{Z}'\mathbf{u}^*}{\sqrt{T}}\right) \\
&= \left(\frac{C/\sqrt{T}\mathbf{Z}'\mathbf{Z}}{\sqrt{T}} + \frac{\mathbf{Z}'\mathbf{v}}{\sqrt{T}} + \frac{\mathbf{v}^{*\prime}\mathbf{Z}}{\sqrt{T}}\right)\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{-1}\left(\frac{\mathbf{Z}'\mathbf{u}^*}{\sqrt{T}}\right) \\
\mathbf{x}^{*\prime} P_Z \mathbf{u}^* | D_T &\xrightarrow{d^*}_p (Q_{ZZ}C + \mathbf{U} + W^*_{Zu})' Q^{-1}_{ZZ} W^*_{Zu}\Big| \mathbf{U}.
\end{aligned}
\tag{3.23}
$$

Combining two results, using (3.23), we obtain the following representation:

$$
\begin{aligned}
(\hat{\beta}^{*TSLS}_T - \hat{\beta}^{TSLS}_T)|D_T &= (\mathbf{x}^{*\prime} P_Z \mathbf{x}^*)^{-1} (\mathbf{x}^{*\prime} P_Z \mathbf{u}^*) \\
&\xrightarrow{d^*}_d \frac{(\mathbf{Q}_{ZZ}C + \mathbf{U} + W^*_{Zu})'\mathbf{Q}^{-1}_{ZZ} W^*_{Zu}}{(\mathbf{Q}_{ZZ}C + \mathbf{U} + W^*_{Zu})'\mathbf{Q}^{-1}_{ZZ}(\mathbf{Q}_{ZZ}C + \mathbf{U} + W^*_{Zu})}.\Big|\mathbf{U}.
\end{aligned}
$$

In the special unidentified case of $C = \boldsymbol{\pi} = \mathbf{0}$:

$$
(\hat{\beta}^{*TSLS}_T - \hat{\beta}^{TSLS}_T)|D_T \xrightarrow{d^*}_d \frac{(W^*_{Zu} + \mathbf{U})' W^*_{Zu}}{(W^*_{Zu} + \mathbf{U})'(W^*_{Zu} + \mathbf{U})'}.\Big|\mathbf{U}
$$

Fixed Regressor iid Bootstrap

Let the bootstrapped IV estimator based on fixed regressor iid bootstrap in Section 3.9 (just identified case):

$$\begin{aligned}\hat{\beta}_T^{IV*} &= \frac{\sum_{t=1}^T z_t y_t^*}{\sum_{t=1}^T z_t x_t^*} = \frac{\sum_{t=1}^T z_t(\hat{\beta}^{IV} x_t + u_t^*)}{\sum_{t=1}^T z_t y_t^*} \\ &= \frac{\hat{\beta}_T^{IV} \sum_{t=1}^T z_t x_t}{\sum_{t=1}^T z_t x_t} + \frac{\sum_{t=1}^T z_t u_t^\star}{\sum_{t=1}^T z_t x_t} \\ &= \hat{\beta}_T^{IV} + \frac{\sum_{t=1}^T z_t u_t^\star}{\sum_{t=1}^T z_t x_t}.\end{aligned}$$

Recalling that $u_t^\star \sim N(0, \hat{\sigma}_u^2)$ and assumptions of Section 2.1.2, the bootstrap IV estimator converges to:

$$\sqrt{T}(\hat{\beta}_T^{IV*} - \hat{\beta}_T^{IV}) \xrightarrow{d^*}_p N(0, \hat{\omega}^2)$$

since $\hat{\omega}^{2*} \xrightarrow{p} \hat{\sigma}_u^2 \frac{\sum_{t=1}^T z_t^2}{\sum_{t=1}^T (z_t x_t)^2} = \hat{\omega}^2$, even if $\hat{\omega}^2 \nrightarrow_p \omega^2$.

Remark: inconsintency of variance estimator

The assumption of consistency for the estimator $\hat{\mathbf{\Sigma}}$ seems more unrealistic under poorly relevant instrument. This assumption may be relaxed if we assume that $\hat{\mathbf{\Sigma}}$ converges in distribution to a random matrix $\mathcal{M}$ rather than constant and positive definite Σ . In order to give an intuition of this fact, under weak instrument asymptotics we have inconsistency of variance of u_t:

$$\hat{\sigma}_u^2 \xrightarrow{d} \Psi \tag{3.24}$$

and Ψ is a random variable equal to:

$$\Psi = \sigma_u^2 - 2\sigma_{uv} D + \sigma_v^2 D^2.$$

where $D = w_{zu}/(c^2\sigma_z + w_{zv})$ is the limiting distribution of $(\hat{\beta}_T^{IV} - \beta)$. The main concept regards that both the numerator and denominator of bootstrapped IV estimator presents two components, either normally distributed, conditionally on the original data D_T. Hence, we point out that our asymptotic results regarding *randomness* of the bootstrapped distribution, formally proved in Theorem 1, 2 3 and 4, may be improved considering $\hat{\Sigma} \xrightarrow{d} \mathcal{M}$.

Three R code for bootstrap in IV/TSLS

Here we present three R functions useful in the context of inference in IV/TSLS context. The first concerns boostrapped distribution of TSLS or IV estimator and associated t–statistic (with no exogenous covariates) using (unrestricted) Residual bootstrap with both resampling and fixed instruments. Standardized bootstrapped estimators are also included in the output. The second code regards bias–corrected TSLS and LIML estimator with both pair and residual–based bootstrap (only with fixed instrument). The last function contains restricted residual bootstrap for the t–statistic associated to TSLS estimator, applying methods introduced by Davidson and MacKinnon (2010). To our knowledge, these procedures are not even implemented in R packages.

1. Unrestricted Residual Bootstrap

```
boot_iv_fun<-function(B,data,beta0){
## tsls/iv estimation:
k<-ncol(data)-2
iv <-ivreg(data[,1]~data[,2]|data[,3:(k+2)]);
tsls<-iv$coefficients[2];
ols<-lm(data[,2]~data[,3:(k+2)])
sdt<-summary(iv)$coefficients[2,2]
tstat<-(tsls-beta0)/sdt
t<-nrow(data)
z<-data[,(3:(k+2))]
## Inizializing vectors:
betaB_UR_res<-numeric();t_boot_UR_res<-numeric()
betaB_UR_fix<-numeric();t_boot_UR_fix<-numeric()
## residuals
uvz<-cbind(residuals(iv),residuals(ols),z)
for(j in 1:B){

indices<-sample(1:t,t,replace=TRUE)
ub<-uvz[indices,1];vb<-uvz[indices,2];
```

```
zb<-uvz[indices,3:(k+2)]
## DGP fixed z
if(k==1){xb<-z*ols$coef[2]+vb} else {xb<-z%*%ols$coef[2:(k+1)]+vb}
yb<-tsls*xb+ub;
ivb_UR_fix<-ivreg(yb~xb|z);
## Bootstrapped estimator and t-stat
betaB_UR_fix[j]<-ivb_UR_fix$coefficients[2];
t_boot_UR_fix[j]<-(betaB_UR_fix[j]-tsls)/summary(ivb_UR_fix)$coefficients

## DGP resampled z:
if(k==1){xb2<-zb*ols$coef[2]+vb} else {xb2<-zb%*%ols$coef[2:(k+1)]+vb}
yb2<-tsls*xb2+ub
ivb_UR_res<-ivreg(yb2~xb2|zb)

## Bootstrapped estimator and t-stat:
betaB_UR_res[j]<-ivb_UR_res$coefficients[2];
t_boot_UR_res[j]<-(betaB_UR_res[j]-tsls)/summary(ivb_UR_res)$coefficients

}
### Standardized estimators
betaB_UR1_st<-(betaB_UR_res-tsls)/sdt
betaB_UR2_st<-(betaB_UR_fix-tsls)/sdt
## Save quantities from IV estimation
iv_est<-list(tsls,tstat)

boot<-list(betaB_UR_res,betaB_UR_fix,
betaB_UR1_st,betaB_UR2_st,
t_boot_UR_res,t_boot_UR_fix)



names(boot)<-c('betaB_UR_res','betaB_UR_fix',
'betaB_URres_st','betaB_URfix_st',
't_boot_UR_res",'t_boot_UR_fix')
names(iv_est)<-c('tsls','tstat')
```

```
OUT<-list(boot,iv_est);names(OUT)<-c('boot','iv_est')
return(OUT)
}

## Example dataset with 2 instruments:
rho<-0.1; mu2<-20; t<-100;Sigma<-matrix(c(1,rho,rho,1),2,2)
k<-2;uv <- mvrnorm(t,c(0,0),Sigma);u<-uv[,1];v<-uv[,2];
i_k=matrix(1,nrow=k,ncol=1)
if(mu2==0) pi<-i_k*0 else pi<-i_k%*%sqrt(mu2)/sqrt(t)
if(mu2==0) sqad<-NA else sqad<-sqrt(solve(t(pi)%*%pi)/(t))
## DGP:
beta<-0; I<-diag(1,k)
z <- mvrnorm(t, rep(0,k) ,I)
if(k==1)x <-z*pi+v else{x <-z%*%pi+v}
y <-beta*x+u
data<-cbind(y,x,z)
## Function (199 bootstrap replication)
boot_out<-boot_iv_fun(199,data,0)

## Bootstrap p.values for t-stat
p_value_boot1<-mean(abs(boot_out$boot$t_boot_UR_res)>
abs(boot_out$iv_est$tstat))
p_value_boot2<-mean(abs(boot_out$boot$t_boot_UR_fix)>
abs(boot_out$iv_est$tstat))
p_value_boot1;p_value_boot2

## Card data without exogenous regressors
data<-cbind(card.data$lwage,card.data$educ,card.data$nearc2)
## Function
B<-999
boot_out<-boot_iv_fun(B,data,0)

## Bootstrap p.values for t-stat
p_value_boot1<-mean(abs(boot_out$boot$t_boot_UR_res)>
```

```
abs(boot_out$iv_est$tstat))
p_value_boot2<-mean(abs(boot_out$boot$t_boot_UR_fix)>
abs(boot_out$iv_est$tstat))
## p_value bootstrap
p_value_boot1
p_value_boot2
```

2. Bootstrap–based bias correction

```
boot_bcor<-function(B,data){
k<-ncol(data)-2
## Estimation, tsls and LIML
y<-data[,1];x<-data[,2]
z<-data[,(3:(k+2))]
iv<-ivreg(y~x|z)
ivL<-suppressWarnings(y,x,z);
tsls<-iv$coef[2];liml<-ivL$LIML$point.est
## Fist stage
ols<-lm(data[,2]~data[,3:(k+2)])
t<-nrow(data)

## Inizializing vectors:
betaB_liml_pair<-numeric();betaB_liml_res<-numeric()
betaB_tsls_pair<-numeric();betaB_tsls_res<-numeric()
## Resampled Variables (pair)
yxz<-cbind(y,x,z)
## Residuals
resliml<-(y-x*as.numeric(liml))-mean(y-x*as.numeric(liml))
uvzL<-cbind(resliml,residuals(ols))
uvz<-cbind(residuals(iv),residuals(ols))
for(i in 1:B){
indices<-sample(1:t,t,replace=TRUE)
## DGP PAIR bootstrap
yb<-yxz[indices,1];xb<-yxz[indices,2];
```

```
zb<-yxz[indices,3:(k+2)]
ivbL<-suppressWarnings(ivmodel(yb,xb,zb));
ivb<-ivreg(yb~xb|zb)
 betaB_tsls_pair[i]<-ivb$coef[2]
 betaB_liml_pair[i]<-as.numeric(ivbL$LIML$point.est)
##DGP Residual Bootstrap
ub<-uvz[indices,1];vb<-uvz[indices,2];
ubL<-uvzL[indices,1];
 if(k==1){xb<-z*ols$coef[2]+vb} else {xb<-z%*%ols$coef[2:(k+1)]+vb}
ybR<-tsls*xb+ub;
ybRL<-as.numeric(liml)*xb+ubL
ivbRL<-suppressWarnings(ivmodel(ybR,xb,z));
ivbR<-ivreg(ybRL~xb|z)
 betaB_tsls_res[i]<-ivbR$coefficients[2]
 betaB_liml_res[i]<-ivbRL$LIML$point.est
}
## Bias corrected estimator tsls
 betaBcor_tslsP<-as.numeric(2*tsls-mean(betaB_tsls_pair))
 betaBcor_tslsR<-as.numeric(2*tsls-mean(betaB_tsls_res))
 betaBcor_limlP<-as.numeric(2*liml-mean(betaB_liml_pair))
 betaBcor_limlR<-as.numeric(2*liml-mean(betaB_liml_res))
## List
 Tsls<-list(betaBcor_tslsP,betaBcor_tslsR)
 Liml<-list(betaBcor_limlP,betaBcor_limlR)

names(Tsls)<-c('Bcor_tsls1',0Bcor_tsls2')
names(Liml)<-c('Bcor_liml1','Bcor_liml2')

OUT<-list(Tsls,Liml);names(OUT)<-c('Tsls','LIML')
return(OUT)

}

## card example with two instruments and no exogenous covariates
```

```
data<-cbind(card.data$lwage,card.data$educ,
card.data$nearc4,card.data$nearc2)

bootBCOR<-boot_bcor(399,data)
bootBCOR$Tsls;bootBCOR$Liml
```

3. Bootstrapped t–statistic

3.1 Residual Restricted Bootstrap

```
t_boot_RR<-function(B,data,beta0){
## tsls/iv estimation:
k<-ncol(data)-2;t<-nrow(data)
iv <-ivreg(data[,1]~data[,2]|data[,3:(k+2)]);
tsls<-iv$coefficients[2];
ols<-lm(data[,2]~data[,3:(k+2)])
sdt<-summary(iv)$coefficients[2,2]
##t stat and p-value
tstat<-(tsls-beta0)/sdt
pvalue_t<-2*pnorm(-abs(tstat))
z<-data[,(3:(k+2))];  x<-data[,2]
y<-data[,1]
## Inizializing vectors:
betaB_RR<-numeric();
t_bootRR<-numeric()
## restricted residuals RR
resRR<-(y-beta0*x)-mean(y-beta0*x)
uv_RR<-cbind(resRR*(sqrt(t/(t-2))),residuals(ols)*sqrt(t/(t-k)))
for(j in 1:B){
indices<-sample(1:t,t,replace=TRUE)
ub<-uv_RR[indices,1];vb<-uv_RR[indices,2];
## Bootstrap DGP
if(k==1){xb<-z*ols$coef[2]+vb} else {xb<-z%*%ols$coef[2:(k+1)]+vb}
yb<-beta0*xb+ub;
```

```
ivb_RR<-ivreg(yb~xb|z);
## Bootstrapped (restricted) estimator and t-stat
betaB_RR[j]<-ivb_RR$coefficients[2];
t_bootRR[j]<-(betaB_RR[j]-beta0)/summary(ivb_RR)$coefficients[2,2]
}
## p.value bootstrap
pvalue_tB<-mean(abs(t_bootRR)>abs(tstat))
iv_est<-list(tstat,pvalue_t);names(iv_est)<-c('tstat','pvalue_t')
boot<-list(t_bootRR,pvalue_tB)
names(boot)<-c('t_bootRR','pvboot')
boot
OUT<-list(boot,iv_est);names(OUT)<-c('boot','iv.est')
return(OUT)
}
```

3.2 Restricted Efficient Bootstrap (Davidson and MacKinnon, 2010).

```
t_boot_RE<-function(B,data,beta0){
## tsls/iv estimation:
k<-ncol(data)-2

iv <-ivreg(data[,1]~data[,2]|data[,3:(k+2)]);
tsls<-iv$coefficients[2];
## Davidson-Mackinnon Estimator:
ols<-lm(data[,2]~data[,3:(k+2)]+residuals(iv))
sdt<-summary(iv)$coefficients[2,2]

tstat<-(tsls-beta0)/sdt
pvalue_t<-2*pnorm(-abs(tstat))

t<-nrow(data);z<-data[,(3:(k+2))];   x<-data[,2]
y<-data[,1]
## Inizializing vectors:
betaB_RE<-numeric();
```

```
t_bootRE<-numeric()
## restricted residuals RE
resiv<-(y-beta0*x)-mean(y-beta0*x)
##residual first stage
resols<-residuals(ols)+ols$coefficients[k+2]*residuals(iv)
resols<-resRE-mean(resols)
uv_RE<-cbind(resiv*(sqrt(t/(t-2))),resols*sqrt(t/(t-k)))
for(j in 1:B){
indices<-sample(1:t,t,replace=TRUE)
ub<-uv_RE[indices,1];vb<-uv_RE[indices,2];
if(k==1){xb<-z*ols$coef[2]+vb} else {xb<-z%*%ols$coef[2:(k+1)]+vb}
yb<-beta0*xb+ub;
ivb_RE<-ivreg(yb~xb|z);
## Bootstrapped estimator and t-stat
betaB_RE[j]<-ivb_RE$coefficients[2];
t_bootRE[j]<-(betaB_RE[j]-beta0)/summary(ivb_RE)$coefficients[2,2]
}

pvalue_tB<-mean(abs(t_bootRE)>abs(tstat))
boot<-list(t_bootRE,pvalue_tB)
iv_est<-list(tstat,pvalue_t);names(iv_est)<-c('tstat','pvalue_t')
names(boot)<-c('t_bootRE','pvboot')
boot
OUT<-list(boot,iv_est);names(OUT)<-c('boot','iv.est')

return(OUT)
}

### Card data example
data<-cbind(card.data$lwage,card.data$educ,
card.data$nearc2,card.data$nearc4)

B<-999
out_RR<-t_boot_RR(B,data,0)
```

```
out_RE<-t_boot_RE(B,data,0)

out_RR$iv.est$pvalue_t;  out_RR$boot$pvboot
outRE$iv.est$pvalue;out_RE$boot$pvboot
```

T=100

	MC		Residual Bootstrap (resampled z_t)			Residual Bootstrap (fixed z_t)		
μ^2	$\hat{\mu}^2$	$F > 10$	$Me(\hat{\mu}^{*2})$	$\overline{\hat{\mu}^{*2}}$	$F^* > 10$	$Me(\hat{\mu}^{*2})$	$\overline{\hat{\mu}^{*2}}$	$F^* > 10$
0	0.081	0.003	0.565	1.131	0.008	0.565	1.150	0.008
1	1.105	0.019	1.184	2.169	0.018	1.197	2.211	0.023
5	4.949	0.167	5.173	6.157	0.185	5.179	6.208	0.188
10	10.302	0.508	10.590	11.633	0.524	10.613	11.787	0.529
20	20.182	0.884	20.121	21.820	0.882	20.201	22.089	0.902
40	41.260	0.999	41.744	43.533	0.997	42.313	44.017	1.000
60	60.855	0.999	60.400	63.763	1.000	61.379	64.423	1.000

T=1000

	MC		Residual Bootstrap (resampled z_t)			Residual Bootstrap (fixed z_t)		
μ^2	$\hat{\mu}^2$	$F > 10$	$Me(\hat{\mu}^{*2})$	$\overline{\hat{\mu}^{*2}}$	$F^* > 10$	$Me(\hat{\mu}^{*2})$	$\overline{\hat{\mu}^{*2}}$	$F^* > 10$
0	-0.027	0.002	0.442	0.974	0.008	0.476	0.982	0.008
1	0.911	0.008	1.106	1.911	0.018	1.117	1.916	0.018
5	5.011	0.186	4.981	6.036	0.175	5.021	6.031	0.182
10	10.342	0.516	10.229	11.376	0.511	10.216	11.383	0.513
20	19.693	0.901	19.299	20.746	0.890	19.491	20.777	0.895
40	40.546	0.999	40.503	41.683	1.000	40.276	41.731	1.000
60	59.977	1.000	59.098	61.140	1.000	59.362	61.242	1.000

Table 3.3: Estimated concentration parameter coming from Monte Carlo simulation and residual bootstrap (IV case, one instrument).

Average values

ρ	μ^2/k	$\bar{\tilde{\beta}}_T$	$Me(\tilde{\beta}_T^*)$	$V(\tilde{\beta}_T^*)$	$Skew(\tilde{\beta}_T^*)$	$K(\tilde{\beta}_T^*)$	IQR^*
	0	0.022	0.020	0.863	0.095	18.441	0.844
	1	0.146	0.175	0.988	-0.297	16.491	0.983
	5	0.128	0.198	1.090	-0.575	9.083	1.201
0.5	10	0.092	0.166	1.006	-0.577	4.935	1.247
	20	0.063	0.118	0.982	-0.381	3.630	1.278
	40	0.043	0.083	0.980	-0.261	3.320	1.297
	60	0.038	0.072	0.979	-0.219	3.235	1.303
	0	-0.016	-0.018	0.873	0.062	20.715	0.835
	1	0.334	0.408	0.904	-0.659	20.390	0.879
	5	0.244	0.390	1.007	-1.365	11.754	1.104
0.9	10	0.174	0.302	0.979	-1.027	6.676	1.185
	20	0.114	0.211	0.984	-0.715	4.412	1.257
	40	0.079	0.153	0.977	-0.485	3.608	1.281
	60	0.067	0.131	0.976	-0.400	3.419	1.293

Median values

ρ	μ^2/k	$\bar{\tilde{\beta}}_T$	$Me(\tilde{\beta}_T^*)$	$V(\tilde{\beta}_T^*)$	$Skew(\tilde{\beta}_T^*)$	$K(\tilde{\beta}_T^*)$	IQR^*
	0	0.009	0.003	0.840	0.124	9.487	0.882
	1	0.147	0.192	0.980	-0.259	7.908	1.048
	5	0.112	0.181	0.995	-0.555	4.601	1.201
0.5	10	0.089	0.163	0.974	-0.521	3.826	1.232
	20	0.061	0.119	0.973	-0.359	3.408	1.278
	40	0.048	0.082	0.956	-0.245	3.200	1.287
	60	0.040	0.074	0.959	-0.207	3.155	1.294
	0	0.004	0.004	0.867	-0.054	9.780	0.884
	1	0.365	0.477	0.871	-0.720	10.184	0.929
	5	0.230	0.384	0.915	-1.214	6.423	1.088
0.9	10	0.165	0.291	0.959	-0.945	4.817	1.188
	20	0.116	0.217	0.952	-0.651	3.820	1.240
	40	0.081	0.155	0.951	-0.473	3.443	1.264
	60	0.065	0.125	0.957	-0.377	3.260	1.292

Table 3.4: Bootstrapped standardized TSLS with $k = 3$ under different strength of instruments and degrees of endogeneity, considering $M = 1000$ data–sets.

$T = 100$

ρ	μ^2/k	$\bar{\tilde{\beta}}_T$	$Me(\tilde{\beta}_T^*)$	$V(\tilde{\beta}_T^*)$	$Skew(\tilde{\beta}_T^*)$	$K(\tilde{\beta}_T^*)$	IQR^*
	0	-0.011	-0.013	0.589	0.032	6.276	0.879
	1	0.213	0.235	0.731	-0.213	5.236	1.017
	5	0.239	0.286	0.859	-0.381	3.925	1.176
0.5	10	0.188	0.233	0.909	-0.343	3.560	1.231
	20	0.134	0.173	0.930	-0.247	3.300	1.262
	40	0.100	0.127	0.935	-0.180	3.167	1.277
	60	0.082	0.105	0.945	-0.146	3.153	1.283
	0	-0.017	-0.017	0.581	-0.024	6.046	0.870
	1	0.549	0.605	0.578	-0.572	6.002	0.880
	5	0.460	0.553	0.747	-0.822	5.061	1.068
0.9	10	0.352	0.435	0.820	-0.646	4.113	1.152
	20	0.251	0.322	0.903	-0.474	3.572	1.233
	40	0.182	0.235	0.926	-0.349	3.358	1.260
	60	0.150	0.190	0.931	-0.268	3.221	1.271

T=1000

ρ	μ^2/k	$\bar{\tilde{\beta}}_T$	$Me(\tilde{\beta}_T^*)$	$V(\tilde{\beta}_T^*)$	$Skew(\tilde{\beta}_T^*)$	$K(\tilde{\beta}_T^*)$	IQR^*
	0	-0.003	-0.001	0.596	-0.008	5.914	0.881
	1	0.201	0.224	0.758	-0.204	5.151	1.042
	5	0.244	0.299	0.903	-0.424	3.921	1.210
0.5	10	0.192	0.245	0.943	-0.373	3.527	1.262
	20	0.144	0.187	0.973	-0.278	3.257	1.299
	40	0.106	0.140	0.983	-0.209	3.113	1.319
	60	0.086	0.110	0.987	-0.159	3.066	1.328
	0	0.246	0.263	0.562	-0.110	6.579	0.841
	1	0.480	0.522	0.549	-0.373	4.492	0.899
	5	0.489	0.602	0.746	-0.788	4.197	1.076
0.9	10	0.370	0.467	0.859	-0.687	4.155	1.209
	20	0.281	0.357	0.927	-0.528	3.528	1.246
	40	0.181	0.237	1.017	-0.371	3.285	1.337
	60	0.147	0.223	1.051	-0.308	3.115	1.356

Table 3.5: Bootstrapped standardized TSLS with $k = 5$ under different degrees of endogeneity and two sample sizes. Results refer to median values through $M = 1000$ replications.

LIML

ρ	μ^2/k	$\bar{\tilde{\beta}}_T$	$Me(\tilde{\beta}^*_T)$	$V(\tilde{\beta}^*_T)$	$Skew(\tilde{\beta}^*_T)$	$K(\tilde{\beta}^*_T)$	IQR^*
	0	0.025	0.017	**183.208**	0.398	**175.177**	1.493
	1	-0.040	0.092	174.463	-0.498	**173.367**	1.691
	5	-0.146	0.013	3.456	-0.966	31.589	1.578
0.5	10	-0.096	0.002	1.395	-0.567	4.294	1.458
	20	-0.058	0.001	1.155	-0.349	3.391	1.387
	40	-0.042	-0.002	1.061	-0.232	3.192	1.361
	60	-0.026	0.008	1.029	-0.167	3.128	1.342
	0	0.013	-0.006	**189.701**	-0.195	**191.263**	1.458
	1	-0.262	0.120	139.679	-1.918	**155.218**	1.475
	5	-0.228	0.001	1.691	-1.582	9.103	1.406
0.9	10	-0.143	0.000	1.217	-0.912	4.550	1.362
	20	-0.095	0.000	1.075	-0.601	3.657	1.332
	40	-0.061	0.005	1.029	-0.413	3.345	1.327
	60	-0.049	0.001	1.017	-0.316	3.186	1.334

Fuller

ρ	μ^2/k	$\bar{\tilde{\beta}}_T$	$Me(\tilde{\beta}^*_T)$	$V(\tilde{\beta}^*_T)$	$Skew(\tilde{\beta}^*_T)$	$K(\tilde{\beta}^*_T)$	IQR^*
	0	0.016	0.009	1.173	0.017	3.033	1.458
	1	0.085	0.098	1.768	0.190	4.015	1.545
	5	-0.081	0.013	1.589	-0.434	5.059	1.481
0.5	10	-0.083	-0.001	1.312	-0.500	3.932	1.435
	20	-0.056	-0.005	1.109	-0.324	3.356	1.368
	40	-0.034	0.000	1.051	-0.214	3.158	1.351
	60	-0.032	-0.004	1.017	-0.162	3.100	1.337
	0	0.018	0.004	1.235	0.008	3.040	1.469
	1	0.199	0.165	1.110	0.853	5.447	1.217
	5	-0.158	-0.004	1.168	-0.991	4.825	1.310
0.9	10	-0.131	-0.006	1.120	-0.783	4.154	1.318
	20	-0.094	-0.004	1.058	-0.559	3.560	1.328
	40	-0.066	-0.007	1.027	-0.381	3.301	1.318
	60	-0.050	0.001	1.008	-0.319	3.215	1.321

Table 3.6: Bootstrapped standardized $\kappa-$class estimators with $k = 5$ under two different degrees of endogeneity. The results refer to median values obtained through $M = 1000$ replication.

ρ	μ^2/k	Mean	Median	MAD	RMSE	Coverage	IDR	KS
0.25	1	0.043	0.078	0.414	0.626	0.901	1.213	0.075
	5	0.016	0.018	0.192	0.217	0.941	0.518	0.058
	10	0.007	0.011	0.148	0.151	0.940	0.382	0.046
	20	-0.001	0.002	0.105	0.106	0.940	0.276	0.026
	40	-0.003	-0.004	0.077	0.075	0.929	0.191	0.032
	60	-0.004	-0.005	0.060	0.059	0.937	0.146	0.046
0.5	1	0.107	0.154	0.409	0.611	0.785	1.156	0.152
	5	0.009	0.035	0.199	0.226	0.897	0.532	0.080
	10	-0.007	0.005	0.151	0.153	0.921	0.383	0.031
	20	0.000	0.005	0.100	0.105	0.923	0.268	0.027
	40	0.000	0.002	0.075	0.074	0.929	0.187	0.037
	60	0.000	0.001	0.055	0.058	0.925	0.146	0.024
0.75	1	0.217	0.254	0.338	0.499	**0.564**	0.996	0.270
	5	0.009	0.032	0.203	0.231	0.826	0.520	0.077
	10	0.005	0.012	0.152	0.157	0.856	0.382	0.069
	20	-0.001	0.006	0.098	0.103	0.915	0.256	0.033
	40	0.002	0.003	0.072	0.073	0.923	0.183	0.034
	60	0.003	0.005	0.058	0.058	0.929	0.150	0.036
0.9	1	0.226	0.291	0.310	0.508	**0.390**	0.888	0.289
	5	0.010	0.050	0.202	0.238	0.769	0.539	0.103
	10	0.010	0.034	0.134	0.151	0.851	0.377	0.098
	20	0.001	0.008	0.101	0.107	0.876	0.272	0.049
	40	0.000	0.005	0.069	0.073	0.912	0.181	0.034
	60	0.000	0.004	0.061	0.061	0.923	0.151	0.029

Table 3.7: Performance of bootstrap–based bias corrected TSLS with $T = 100$ running $B = 399$. The number of instruments is $k = 5$

ρ	μ^2/k	Mean	Median	MAD	RMSE	Coverage	IDR	KS
0.25	1	-0.460	0.066	0.842	17.354	0.997	4.501	0.182
	5	0.021	0.011	0.232	3.704	0.996	0.645	0.058
	10	-0.001	0.003	0.149	0.416	0.978	0.387	0.033
	20	0.002	0.003	0.100	0.103	0.961	0.262	0.021
	40	0.001	0.004	0.073	0.076	0.937	0.192	0.026
	60	0.001	0.000	0.058	0.059	0.948	0.144	0.022
0.50	1	1.137	0.153	0.824	42.922	0.988	5.521	0.194
	5	0.077	0.035	0.216	1.608	0.979	0.599	0.086
	10	0.006	0.013	0.138	0.149	0.968	0.350	0.052
	20	0.002	0.004	0.102	0.104	0.957	0.260	0.030
	40	0.000	0.001	0.072	0.073	0.946	0.184	0.021
	60	-0.004	-0.002	0.062	0.061	0.949	0.154	0.035
0.75	1	4.433	0.159	0.726	132.068	0.985	5.125	0.214
	5	0.036	0.063	0.177	2.184	0.969	0.533	0.142
	10	0.001	0.015	0.139	0.148	0.967	0.365	0.063
	20	0.005	0.012	0.094	0.102	0.951	0.257	0.062
	40	0.001	0.006	0.067	0.072	0.949	0.178	0.049
	60	0.000	0.002	0.059	0.057	0.956	0.144	0.025
0.90	1	6.689	0.303	0.566	**186.688**	0.980	4.378	0.263
	5	0.241	0.063	0.168	4.849	0.948	0.465	0.139
	10	-0.002	0.013	0.145	0.146	0.949	0.356	0.044
	20	-0.001	0.010	0.102	0.105	0.948	0.261	0.046
	40	0.002	0.006	0.072	0.073	0.947	0.188	0.046
	60	0.000	0.001	0.051	0.058	0.941	0.148	0.040

Table 3.8: Performance of Bootstrap-Bias corrected LIML with $T = 100$ and $B = 399$. The number of instruments is $k = 5$

IV Case ($k = 1$)

Design		Length			Coverage		
ρ	μ^2/k	Asymptotic	Percentile	t–boot	Asymptotic	Percentile	t–boot
	0	7.059	**23.126**	5.284	0.980	0.993	**0.477**
	1	3.885	**17.120**	2.898	0.966	0.979	0.564
	5	1.847	3.946	1.511	0.965	0.975	0.726
0.5	10	1.273	1.786	1.142	0.957	0.967	0.824
	20	0.881	1.040	0.837	0.955	0.955	0.873
	40	0.623	0.668	0.606	0.963	0.948	0.922
	60	0.504	0.538	0.509	0.959	0.955	0.941
	0	3.585	11.839	2.760	0.704	0.957	**0.329**
	1	3.955	**16.660**	4.444	0.842	0.968	0.503
	5	1.744	**4.557**	1.817	0.900	0.970	0.687
0.9	10	1.198	1.907	1.202	0.927	0.973	0.779
	20	0.870	1.069	0.877	0.947	0.968	0.856
	40	0.626	0.694	0.630	0.951	0.954	0.916
	60	0.503	0.551	0.516	0.948	0.947	0.928

TSLS ($k = 5$)

Design		Length			Coverage		
ρ	μ^2/k	Asymptotics	Percentile	t–boot	Asymptotics	Percentile	t–boot
	0	1.759	1.433	1.556	0.809	0.787	0.562
	1	1.203	1.107	1.121	0.832	0.822	0.727
	5	0.706	0.690	0.708	0.900	0.893	0.864
0.5	10	0.525	0.515	0.535	0.932	0.929	0.917
	20	0.382	0.381	0.393	0.941	0.934	0.928
	40	0.273	0.273	0.280	0.941	0.935	0.947
	60	0.225	0.226	0.232	0.931	0.926	0.933
	0	0.881	0.727	0.794	0.106	**0.001**	0.168
	1	0.839	0.670	**0.965**	0.489	0.220	0.663
	5	0.651	0.581	0.747	0.811	0.757	0.875
0.9	10	0.510	0.479	0.562	0.859	0.838	0.904
	20	0.374	0.360	0.393	0.903	0.893	0.931
	40	0.270	0.264	0.280	0.937	0.932	0.953
	60	0.224	0.223	0.233	0.944	0.939	0.952

Table 3.9: Asymptotic and bootstrapped confidence sets (TSLS, $B = 399$) under different degrees of identification and two levels of endogeneity.

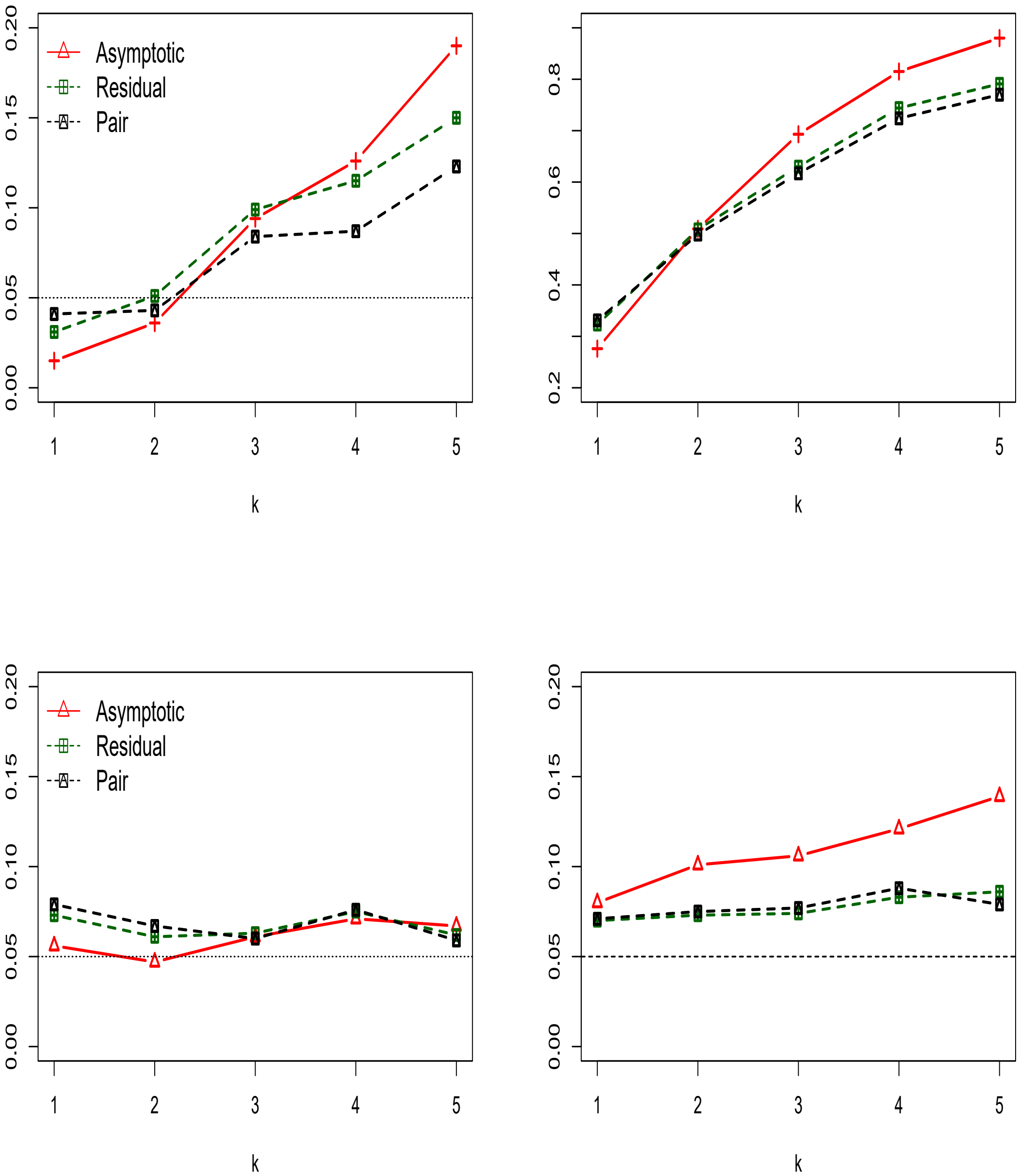
Pair Bootstrap

Design		Length			Coverage		
ρ	μ^2/k	Asymptotic	Percentile	t–boot	Asymptotic	Percentile	t–boot
0.5	0	3.695	**20.497**	4.471	0.901	0.992	**0.376**
	1	1.785	9.896	1.945	0.917	0.993	0.629
	5	0.798	1.157	0.844	0.951	0.977	0.898
	10	0.566	0.670	0.597	0.953	0.949	0.924
	20	0.400	0.426	0.407	0.957	0.946	0.932
	40	0.282	0.287	0.283	0.950	0.944	0.944
	60	0.230	0.233	0.230	0.944	0.931	0.932
0.9	0	1.913	**10.574**	2.311	0.537	0.945	**0.314**
	1	1.796	9.009	2.371	0.888	0.981	0.692
	5	0.797	0.987	0.813	0.927	0.944	0.856
	10	0.561	0.624	0.568	0.953	0.934	0.902
	20	0.397	0.415	0.401	0.953	0.952	0.936
	40	0.282	0.283	0.279	0.936	0.926	0.926
	60	0.229	0.231	0.228	0.955	0.946	0.948

Residual Bootstrap

Design		Length			Coverage		
ρ	μ^2/k	Asymptotics	Percentile	t–boot	Asymptotics	Percentile	t–boot
0.5	0	3.490	10.445	3.621	0.896	0.881	**0.535**
	1	1.712	3.162	1.667	0.930	0.894	0.757
	5	0.797	0.878	0.797	0.943	0.904	0.892
	10	0.561	0.579	0.566	0.946	0.939	0.927
	20	0.402	0.402	0.404	0.941	0.925	0.919
	40	0.281	0.280	0.285	0.956	0.951	0.946
	60	0.230	0.228	0.231	0.950	0.934	0.945
0.9	0	1.748	5.543	1.765	0.489	0.769	**0.378**
	1	1.695	2.506	1.811	0.892	0.873	0.839
	5	0.795	0.854	0.806	0.931	0.917	0.888
	10	0.566	0.581	0.569	0.939	0.921	0.918
	20	0.399	0.399	0.400	0.960	0.941	0.941
	40	0.285	0.282	0.286	0.941	0.938	0.939
	60	0.229	0.226	0.231	0.947	0.941	0.948

Table 3.10: Asymptotic and bootstrapped confidence sets (LIML, B=399) under different degrees of identification and two levels of endogeneity.



(a) $\rho = 0.5$ (b) $\rho = 0.9$

Figure 3.6: Rejection frequencies of bootstrap t–test against number of instruments under irrelevant (upper panel) and strong instruments (lower panel, $\mu^2/k = 0, 10$)

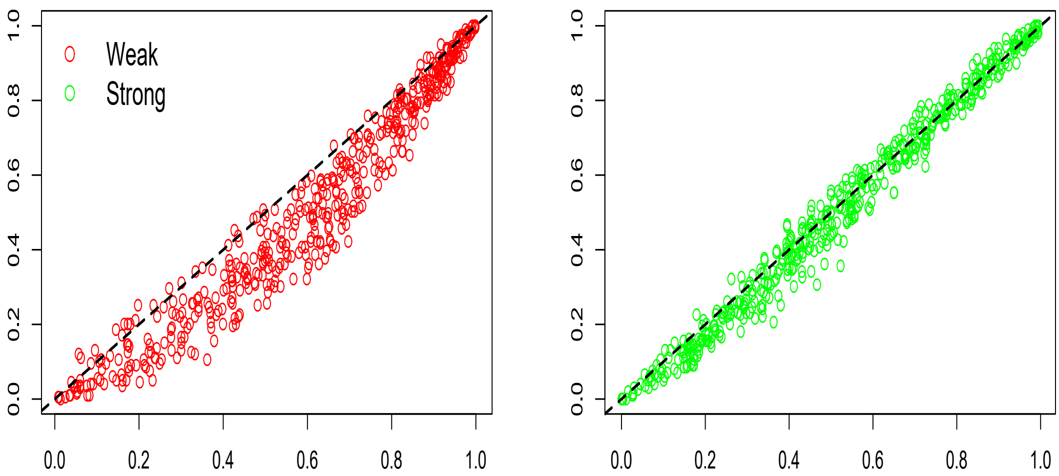


Figure 3.7: Fanchart of asymptotic p–values (x–axis) against bootstrap–based p–values (y–axis) in the just–identified case.

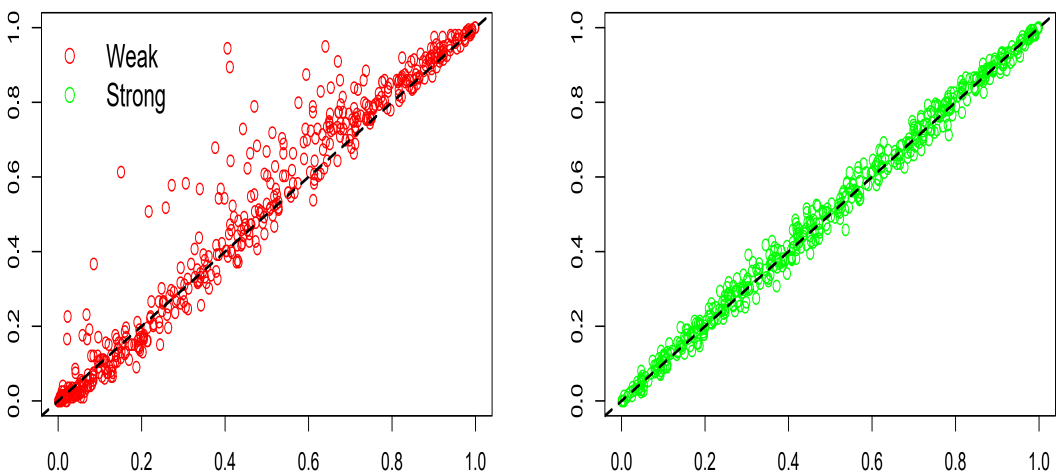


Figure 3.8: Fanchart of asymptotic p–values (x–axis) against bootstrap–based p–values (y–axis) in the overidentified case ($k = 5$).

$k = 1$

ρ	μ^2/k	Asymptotic	Pair	Residual
	0	0.011	0.022	0.029
	1	0.027	0.038	0.049
	5	0.038	0.040	0.074
0.5	10	0.042	0.053	0.063
	20	0.042	0.051	0.054
	40	0.042	0.049	0.058
	60	0.036	0.040	0.055
	0	0.262	0.132	**0.320**
	1	0.167	0.090	0.157
	5	0.111	0.075	0.092
0.9	10	0.064	0.052	0.060
	20	0.074	0.064	0.071
	40	0.059	0.055	0.062
	60	0.046	0.046	0.048

$k = 5$

ρ	μ^2/k	Asymptotic	Pair	Residual
	0	0.196	0.089	0.125
	1	0.142	0.074	0.130
	5	0.082	0.058	0.081
0.5	10	0.075	0.059	0.072
	20	0.060	0.051	0.061
	40	0.046	0.044	0.047
	60	0.064	0.063	0.062
	0	0.888	0.325	**0.794**
	1	0.509	0.204	0.302
	5	0.179	0.079	0.092
0.9	10	0.124	0.076	0.071
	20	0.096	0.051	0.061
	40	0.078	0.058	0.060
	60	0.059	0.052	0.053

Table 3.11: Rejection frequencies of Wald test for the null hypothesis $H_0 : \beta = 0$ associated to IV and TSLS estimation.

$\phi = 0.1$

ρ	μ^2/k	Asymptotic	Pair	Residual
	0	0.010	0.025	0.037
	1	0.033	0.041	0.070
	5	0.049	0.058	0.086
0.5	10	0.038	0.052	0.068
	20	0.054	0.063	0.076
	40	0.048	0.053	0.056
	60	0.050	0.053	0.060
	0	0.240	0.111	**0.295**
	1	0.197	0.107	0.203
	5	0.115	0.084	0.108
0.9	10	0.095	0.083	0.088
	20	0.074	0.063	0.077
	40	0.064	0.064	0.066
	60	0.054	0.054	0.064

$\phi = 1$

ρ	μ^2/k	Asymptotic	Pair	Residual
	0	0.020	0.036	0.032
	1	0.104	0.139	0.139
	5	0.200	0.233	0.293
0.5	10	0.235	0.261	0.318
	20	0.214	0.242	0.272
	40	0.207	0.215	0.232
	60	0.217	0.225	0.233
	0	0.162	0.088	0.131
	1	0.491	0.308	0.531
	5	0.421	0.334	0.406
0.9	10	0.360	0.299	0.339
	20	0.296	0.272	0.291
	40	0.287	0.281	0.298
	60	0.253	0.248	0.260

Table 3.12: Rejection frequencies of Wald test for the null hypothesis $H_0 : \beta = 0$ associated to IV considering a weak and (weakly) endogenous instrument

Chapter 4

Bootstrap–based tests and diagnostic for weak instruments

In this chapter we exploit the possibility of detecting and/or testing weak instruments via bootstrap methods, starting from non-normality of the bootstrapped distribution of IV/TSLS estimators under weak instrument asymptotics, theoretically analized and discussed in Chapter 3. Firstly, we introduce and discuss graphical inspection and descriptive measures, some of which are previously suggested by Zhan (2017)[1]. Since some of these measures strictly depend on the sample size and number of bootstrap replications, we decide to develop new bootstrap–based tests to for the relevance of instruments taking account of other IV/TSLS issues as the no–moment problem, amount of overidentification and levels of endogeneity. The method basically relies on the work of Angelini et al. (2016), introducing bootstrap methods in the frequentist evaluation of (dynamic) rational expectation models, here adapted for the first time in the linear IV framework and $\kappa-$class estimation. The key idea concerns the difference between the unknown cumulative distribution of a bootstrapped statistic of interest, here denoted $\tilde{\beta}_T^*$, with its limit distribution, under the assumption of strong identification and if other regularity conditions for inference holds, e.g. validity of instruments. Practically speaking, this procedure allows to apply conventional normality tests directly on the (possibly normalized) B bootstrapped replications of a given estimator, representing a formal test of asymptotic normality.

However, performancs in terms of rejection frequencies could be affected by three main issues: the first is lack of moments for IV estimator, occurring in perfectly iden-

[1]Unpublished paper

tified case. We empirically observe this problem in huge RMSE measures of Section 2.3 considering several $\kappa-$class estimators. The second regards presence of high endogeneity ($\rho \approx 1$), because it affects finite sample behaviour of estimators, as mentioned in Chapter 2. Finally, when the number of instruments increases, performance of conventional normality tests may be unsatisfactory, especially in terms of power. From a different point of view, a similar lack of power may be found in the test proposed by Hahn and Hausman (2002a), which presents the null hypothesis of strong/valid instruments. Sources of poor rejection frequencies may be twofold. Firstly, the random component of the conditional distribution, defined as $\mathbf{U}$, plays a central role in this sense, as pointed out in Section 3.2.3. Secondly, TSLS estimator is asymptotically normal under the so–called many instrument sequence, introduced by Bekker (1994), even if it is not consistent, and the bootstrap is also invalid in this context (see Wang and Kaffo, 2016). Anyway, on average, the differences between asymptotic distribution of $\hat{\beta}_T^{TSLS}$ and its (random) bootstrap counterpart affects the distribution of $\tilde{\beta}_T^*$, as shown in simulation of Section 3.3, in terms of mean, variance and other sampled moments. In order to overcome these issues, we propose to test asymptotic normality with zero mean, or alternately, standard normality of $\tilde{\beta}_T^*$, where the null hypothesis is $H_0 : \tilde{\beta}_T^* \sim N(0, 1)$. This may be done applying a modified Shapiro–Wilk statistic, and then proposing new test statistics based on the sample moments of standard normal. At the end of this Chapter we apply new bootstrap based tests on two real well–known datasets in instrumental variable literature, to show that these methods may help to detect weak instruments.

4.1 Bootstrap and diagnostic

When inference is conducted through commonly used estimators, i.e. IV and TSLS, the bootstrap is deemed not valid, considering the standard definition of validity in a bootstrap sense, when the instruments are collectively weak. This means that it does not provide a first-order improvement, giving a poor approximation of asymptotic distributions of estmators and test statistics. As shown by Moreira et al. (2009), this failure is related to a violation in the regularity conditions occurring in the Edgeworth expansions. This appears in the bootstrap version of t–test under weak instruments, previously exposed in Section 3.2.6 and whose failures are shown in simulation study of Section 3.3. However, validity of bootstrap can be restored applying statistics that are robust to weak instruments, see for example Moreira (2003). Anyway, in this context, application of bootstrap methods is not helpful to provide information on the identification level.

Furthermore, when instruments are not deemed weak, the bootstrap distribution of IV estimator is asymptotically normal and so the standardized distribution of IV estimator can be close to the standard normal, as we point out in Section 3.1. As mentioned, bootstrap distribution of IV estimator $(\hat{\beta}^*_{T1}, \ldots, \hat{\beta}^*_{TB})$ and the associate t/Wald test, given the sample, preserves some features of the (unknown) asymptotic distribution of the estimator, although it present some random components. In particular, when the number of instruments is equal to the explanatory endogenous variables, the bootstrap distribution of the estimators could be dramatically non–normal under weak identification, with huge standard errors and heavy or sometimes tails, as shown in simulation of Section 3.3. This is due to the combination of poorly relevant instruments with the non existence of moments in the perfectly identified cases[2] and, in general, when $k < 3$, as proved by Mariano (2001). This features affect the distribution of an estimator and its bootstrap counterpart, meaning that it has very huge variability and unreliable dispersion under weak identification, especially in small samples. For this purpose, detection tools to evaluate non–normality of bootstrap estimators could be the following: a) graphical inspection, b) Kolmogorov Smirnov Distance, c) Bootstrap Mean square error.

4.1.1 Graphical Inspection

A first check of the strength of identification could be conducted through graphical inspections, comparing the empirical density of the bootstrap replications: $(\hat{\beta}^*_{T1}, \ldots, \hat{\beta}^*_{TB})$ against its theoretical limiting distribution, conditionally on the data, i.e. $N(\hat{\beta}_T, \hat{V}(\hat{\beta}_T))$, where $\hat{\beta}_T$ is one of the discussed limited information estimators and $\hat{V}(\hat{\beta}_T)$ stand for its estimated variance.

As viewed in simulation of Section 3.3, it is possible to standardize the bootstrap distribution of the parameters and evaluate distance from the standard normal in terms of mean, median and variance. Loosely speaking, it could be useful to check if mean and median of the standardized bootstrap distribution are close to zero or if the variance is higher or lower than 1.

[2] This occurs even in the LIML case

To summarize, malfunction in conventional asymptotic can be viewed as an evidence of weak instruments. For this kind of comparison, pdf plots, Ecdf plots, Q–Q Plot and P–P Plot may be useful to verify the presence of heavy tails and high level of kurtosis, which appears when very weak identification is combined with the lack of moments, as pointed out in the simulation of Section 2.3. Boxplots of the bootstrapped distribution also can help to detect the presence of outliers and could be used to compare a) different bootstrap estimators, included in the $\kappa-$class b) different resampling scheme, previously discussed in Section 2.4.

4.1.2 Bootstrapped Kolmogorov Smirnov distance

Bootstrapped KS statistic is suggested by Zhan (2017) as a descriptive measure of weak identification in IV/TSLS models. From a practical perspective, KS represents the worst–case size distortion for the Wald/t-test statistic using the critical values from the normal distribution, since $\tilde{\beta}_T$ is the non-studentized statistic, equal to $T^{1/2}\omega^{-1}(\hat{\beta}-\beta)$. In Chapter 2 we show the desirable behaviour of KS statistic in the evaluation of finite sample distribution of $\kappa-$class estimators, through Monte Carlo simulations. We also propose this method in order to compare bias–correted TSLS/LIML under weak instruments in section 3.3, where the amount of bias is estimated through bootstrap methods. To summarize, low values of KS suggest the use of conventional TSLS/IV inference, whereas when KS increases, the standardized TSLS distribution could be far from the standard normal.

Since the true maximal difference between Cdfs is not known, the bootstrap is proposed to estimate the unknown quantity of (2.31) where its counterpart, denoted by KS^*, is defined as follows:

$$KS^* = \sup_{-\infty<c<+\infty} \left| P\left(\tilde{\beta}^*_T \leq c \middle| (\mathbf{y}, \mathbf{x}, \mathbf{Z})\right) - \Phi(c) \right|,$$

and measures the distance between the bootstrap Ecdf of β^{*IV} and its asymptotic Cdf from the $N(\hat{\beta}, \hat{\sigma}_{\hat{\beta}})$, where $\hat{\beta}$ and $\hat{\sigma}_{\hat{\beta}}$ are the TSLS/IV estimates from the data. From a different point of view, KS^* cold be thought a descriptive measure of how the G^* of any bootstrapped statistic is far from G_∞, i.e. the cdf of τ_∞. In order to give an empirical threshold, instruments may deemed collectively weak if the KS^* exceeds a given threshold (e.g. 0.05 or 0.10).

Zhan demonstrate (Theorem 3) that KS and its bootstrap counterpart share the same order in probability $O_p\left(T^{-1/2}\right)$ when identification is deemed strong and KS^* is also a super–consistent estimator for KS. In order to build a formal bootstrap test, a computationally demanding double[3] bootstrap Horowitz (2001) is implemented to obtain the unknown distribution of KS^*, bootstrapping the residual from bootstrap samples $D_T^* = (y_t^*, x_t^*, Z_t^*)$ and computing $KS_1^{**}, \ldots, KS_B^{**}$, where the double subscript. The test is constructed estimating the α quantile of KS^* with KS_α^{**} using *percentile* bootstrap methods. The null hypothesis of strong identification is rejected when $KS_\alpha^{**} \leq \alpha$.

However, both KS^* as descriptive indicator and proposed bootstrap–based test suffer from two main limitation. Firstly, the KS^* requires moderately large number of replications B, and then bootstrap–based test could be computationally demanding because it needs B^2 iterations. In addition, KS test seems to reach desired rejection frequencies, in terms of empirical power under weak instruments especially when endogeneity is unrealistically very close to one, in just identified case. Nonetheless, it could be investigate if KS^* may be a useful tool to evaluate the failures of bootstrap in inference conducted through t–statistic, (using τ_T^*), since its asymptotic distribution, in large samples, is standard normal.

4.1.3 Bootstrap mean square error

As mentioned, the difference between $\hat{\beta}_T^*$ and $\hat{\beta}_T$ could be thought as a diagnostic of weak instruments, recalling the fact that both converge to a non–normal distribution under this particular modeling, differing from a random component $\mathbf{U}$. This difference could be exacerbated if the instruments are very weak, as already pointed out in examples of Section 3.2.3. In this context the Bootstrap Mean Square Error, defined as MSE^* may be used as an estimates of the MSE in TSLS/IV estimation:

$$MSE^* = B^{-1} \sum_{b=1}^{B} (\hat{\beta}_T^* - \hat{\beta}_T)^2. \tag{4.1}$$

Since it reflects the variability around $\hat{\beta}_T$, MSE^* could be used as a descriptive measure of weak instruments. We show that, in the just identified case, MSE^* can take huge values, due to the lack of moments, while it rapidely decreases with number of overidentification. Furthermore, for a given k, the MSE^* simulation evidences, ap-

[3]Iterated two times i.e. involving two loops

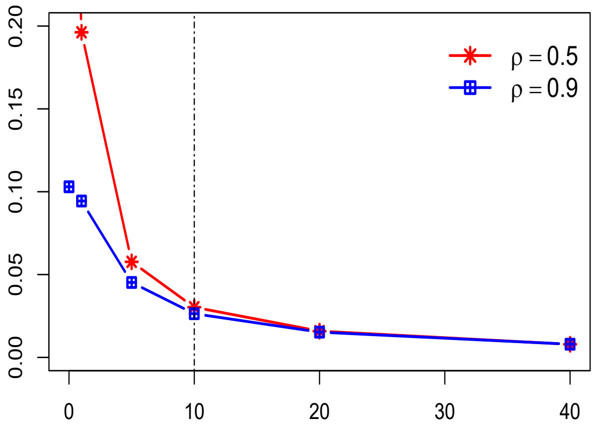


Figure 4.1: Median of MSE^* among different levels of identification $(k = 5)$

plying weak instrument asymptotics, suggest that it decreases with the strength of identification, represented by μ^2/k. However, to our knowledge it is not possible to choose a proper threshold, given the number of instruments, which allows to distinguish between strong and weak instruments.

Figure 4.1 shows the results of a small–scale Monte Carlo exercise under $k = 3$, $B = 199$ and $M = 1000$ replication. We plot the median of MSE^* for a small–scale exercise with TSLS against the identification level where $k = 5$ and using two levels of endogeneity (red line stand for moderately ρ and blue line for $\rho = 0.9$). In this case, bootstrapped MSE rapidly decreases when the strength of instruments increases, resulting lower in case of high endogeneity. Although beyond the purpose of this work, it seems that MSE^* can be used in the choice of instrument, following the approach of Donald and Newey (2001) and applying bootstrap as suggested by Wang et al. (2015).

4.2 A new bootstrap–based test

In this section we present the methodology of a novel bootstrap–based misspecification test for asymptotic normality of a bootstrap statistic or estimator. As above mentioned, weak instruments, combined with lack of moments, leads to non–normality of TSLS/IV estimators in finite samples and these features are also reflected in their bootstrap counterpart, as proved in Section 3.2. In addition, the randomn component of bootstrapped distribution could be exacerbated under very weak or irrelevant instruments. Nevertheless, applying normality tests on B replications does not seem a reasonable choice, because the distribution of bootstrapped estimator (or test statistic) is gaussian only in the limit, as $B, T \to \infty$, even if B may be moderately large in bootstrap procedure. So, applying previously introduced bootstrap methods, normality tests tend to reject too often, even if the instruments may be collectively very strong.

In this sense, Angelini et al. (2016) introduce a bootstrap–based misspecification test in the context of estimation and evaluation of dynamic rational expectation models, where identification issues are still under debate both in the frequentist framework (see Canova and Sala, 2009). The basic idea is to apply normality tests on a certain number of bootstrap estimates, in order to test how the bootstrapped distribution of structural parameters is far from standard asymptotic theory.

In the following sections we firstly discuss the methodology and summarize the procedure of bootstrap–based test. Successively we introduce standard normality tests useful to verify malfunctions in conventional standardization.

4.2.1 Method and asymptotics

The main approach is based on the evaluation of G_T^*, i.e. the cumulative distribution function of a (possibly normalized) bootstrap statistic, here denoted with $\hat{\beta}_T^*$ or $\tilde{\beta}_T^*$. The key point is to evaluate the distance between $G_T^*(\cdot)$ and the cumulative distribution function of standard normal, defined as $G_T^* - \Phi_Z(x)$. Since G_T^* is unknown, it could be estimated from its finite sample (bootstrapped) counterpart, based on B replications, as follows:

$$G_{T,B}^*(x) = B^{-1} \sum_{b=1}^{B} \mathbf{I}(\tilde{\beta}_T^* \leq x), \tag{4.2}$$

where $\mathbf{I}$ stands for the indicator function and expression in (4.2) is the cumulative distribution function of $\left\{\tilde{\beta}_{T1}^*, \ldots \tilde{\beta}_{TB}^*\right\}$. The estimation error $G_{T,B}^*(x) - G_T^*(x)$, for any fixed $x \in \mathbf{R}$, may be properly normalized as follows:

$$B^{1/2} V(x)^{-1/2} (G_{T,B}^*(x) - G_T^*(x)) \xrightarrow{d} N(0,1). \tag{4.3}$$

This equality holds by the CLT when $B \to \infty$. Thus, the distance previously introduced could be estimated as $G^*_{T,B} - \Phi_Z(x)$ and normalized in the following way:

$$d_{T,B} = B^{1/2} V_{T,B}(x)(G^*_{T,B}(x) - \Phi(x)), \tag{4.4}$$

where $V_{T,B}(x)$ is an available consistent estimator of $V_T(x)$, denoting the variance of $G^*_T(x)$. A proper choice could be represented by $G^*_{T,B}(x)(1 - G^*_{T,B}(x))$. Decomposing the expression in (4.4) it is possible to obtain:

$$\begin{aligned} d_{T,B} &= B^{1/2}\hat{V}_{T,B}(x)(G^*_{T,B}(x) - \Phi(x)) \\ &= B^{1/2}\hat{V}_{T,B}(x)(G^*_{T,B}(x) - G_{T,B}(x)) + B^{1/2}\hat{V}_{T,B}(x)(G^*_{T,B}(x) - \Phi(x)). \end{aligned} \tag{4.5}$$

Since the first term is asymptotically standard normal, the second vanishes asymptotically to zero if the following condition holds: a) it admits, under regularity conditions (i.e. strong identification), an Edgeworth Expansion such that $G(x)^*_T = \Phi(x) + O_p(T^{-1/2})$, where $\Phi(x)$ is the Cdf of standard normal, b) the expression (4.4) is bounded in probability: $d_{T,B} = O_p(T^{-1/2}B^{1/2})$, and converges in probability to zero. Briefly speaking, in order to apply the following bootstrap–based test, introduced asymptotics requires that quantity $\overline{B}T^{-1}$ is close to zero, and therefore the choice of $\overline{B}$ may affects performance of bootstrap–based tests in finite samples.

To summarizem this framework allows to apply normality test directly on a finite number of bootstrap replications, $(\hat{\beta}^*_{T,1}, \ldots, \hat{\beta}^*_{T,\overline{B}})$ where $\overline{B}$ is sample length, and then reject null hypothesis if the resulted p–value $p < \alpha$ where α is the chosen nominal level. From another perspective, well–known test statistics, e.g. the Jarque–Bera statistic, have an only asymptotically known free parameters distribution under the null hypothesis of normality, e.g. χ^2 with 2 degrees of freedom. For this reason $\overline{B}$ should not be very small in order to control the size of test. Tentative thresholds, strictly depending on the sample size, may be proposed in order to control empirical size and power. Our preliminary results (not presented here) suggests that JB test performs poorly with $\overline{B} \leq 20$.

One of the potential drawback is represented by the loss of information given by $B - \overline{B}$ bootstrap estimates. In order to overcome this problem, it is possible to apply two procedures involving total number of B replications, dividing the bootstrap sample in N independent non–overlapping groups. The simultaneous test produces N associated p–values $(p_1, \ldots, p_N)$ and, two possible strategy could be adopted for a given significance level α. The former relies on Bonferroni correction, rejecting the null hypothesis if $min_{i=1,\dots,N}(p_i) \leq \alpha_0$, where $\alpha_0 = \alpha/N$, while the latter technique is also based on the stochastically independence between the N tests. In fact, it is possible to select the overall I type error probability as $\alpha = 1 - (1 - \alpha_0)^N$, and again reject normality if $min_{i=1,\dots,N}(p_i) < \alpha_0$ where α_0 is obtained from $\alpha_0 = 1 - (1 - \alpha)^{1/N}$. These procedures may be implemented in different type of bootstrap algorithms, where B can be arbitrary large (usually 399 or 999), even if the bootstrap is applied as a standard tool for inference, e.g. for bias-correction method or to obtain standard error and/or confidence intervals. Different normality tests may be used in this framework: we apply Shapiro–Wilk and, as above mentioned, well–known Jarque–Bera.

However, when the number of instruments increases, the distribution of TSLS estimator could be asymptotically normal (Bekker, 1994), but far from its limiting distribution, i.e. $N(\beta, V(\hat{\beta}_T))$. Furthermore, Monte Carlo simulation of Section 2.3 shows that the standard errors of bootstrapped TSLS are often closer to the true value in overidientified models, even if instruments are collectively weak. In order to control the distance in mean from the standard normal distribution, a Shapiro–Wilk type statistic with *known mean*, introduced by Hanusz et al. (2016), is proposed as an alternative to already mentioned tests. Given that the known mean of $\tilde{\beta}^*$ is $\mu_0 = 0$, the test statistic, denoted as W_0, could be obtained as follows:

$$W_0 = W \cdot \frac{\sum_{b=1}^{\overline{B}} \left(\tilde{\beta}^*_{T,b} - \overline{\tilde{\beta}}^*_{T,\overline{B}} \right)^2}{\sum_{b=1}^{\tilde{B}} \left(\tilde{\beta}^*_{T,b} \right)^2}, \tag{4.6}$$

where W is the Shapiro–Wilk Statistic and $\overline{\tilde{\beta}}^*_{T,\overline{B}}$ is the mean of $(\tilde{\beta}^*_T)$ bootstrap replications, i.e. $B^{-1} \sum_{b=1}^{B} \tilde{\beta}^*_{T,b}$. Thus, the null hypothesis is $H_0 : \tilde{\beta}^*_T \sim N(0, \hat{V}(\hat{\beta}_T))$. Since the statistic in (4.6) has a proper known distribution only if $T = 3$, W_0 could be normalized through the Johnson procedure in the following way:

$$z = \gamma + \delta \tilde{W},$$

where the normalizing constants γ and δ can be substituted by $\hat{\gamma}$ and $\hat{\delta}$, estimated through an OLS regression of z on $\tilde{W} = log(W_0(1 - W_0)^{-1})$. This procedure is implemented in order to obtain p–values of proposed test, computing observed statistic z and then $\Phi(z)$, where Φ denotes the probability of standard normal distribution;

the estimates of δ and γ are found in Hanusz et al. (2016). This procedure presents a possible drawback, because coefficients are estimated using a moderate number of observation ($T = 3, \dots, 50$). Thus, choosing $T > 50$, asymptotic standard normality of z is not guarantee. For this reason we apply this test, denoted as SW_0, using only a small number of bootstrap replication, e.g. $T = \overline{B} = 30, 40, 50$.

Figure 4.2 shows rejection frequencies of proposed Shapiro–Wilk with known mean statistic, obtained through a straigthforward simulation exercise (without bootstrapping). The data are generated from a random variable $N(\tau_i, 1)$, where $\tau_i \in [-1, 1)$, with sample size equal to $T = 20, 30, 40$, and the number of simulations is equal to $M = 100000$. As it clearly appears from the figure, rejection frequencies increase when the expected values deviates from zero, and test presents an high empirical power when $\tau_i > |0.5|$ and the sample size is $T = 40$.

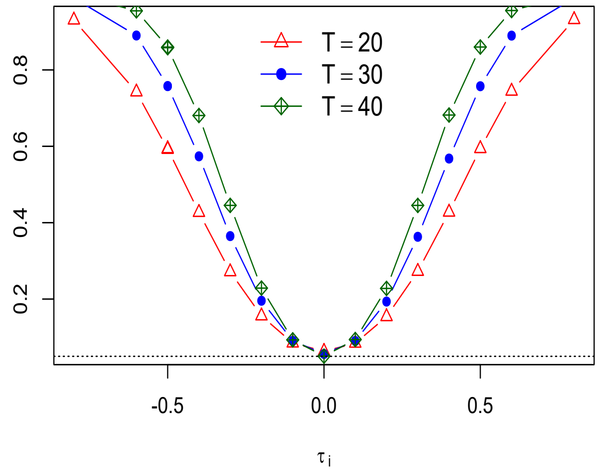


Figure 4.2: Rejection frequency of statistic W_0 for different T. DGP: $X_i \sim N(\tau_i, 1)$

4.2.2 Algorithm for the procedure

In this subsection we present the algorithm to implement proposed test for the asymptotic normality of bootstrap estimator, applying methodology presented in Section 4.2.1. The procedure is summarized in the following steps:

1. Given the sample $D_T = (\mathbf{y}, \mathbf{x}, \mathbf{Z})$, estimate parameters and obtain $\hat{\beta}_T, \hat{\pi}_T$, implied

residuals $(\hat{u}_t, \hat{v}_t)$ and $\hat{\sigma}_u^2$, where $\hat{\beta}_T$ is one of the considered $\kappa-$class estimators and $\hat{\pi}_T$ is estimated through OLS.

2. Apply residual–based resampling scheme, introduced in section 2.4.3, to generate bootstrap sample of endogenous variables[4] $(y_t^*, x_t^*)'$ and instruments Z_t^* or Z_t (not resampled).

3. Use bootstrap sample D_T^* to obtain the bootstrap counterpart of the estimator $\hat{\beta}_T^*$ and its standardized version $\tilde{\beta}_T^* = \sqrt{T}\hat{\omega}^{-1}(\hat{\beta}_T^* - \hat{\beta}_T)$.

4. Repeat steps 1–3 B times, to have bootstrap distribution of estimator $\left\{\tilde{\beta}_{T,b}^*\right\}_{b=1}^{B}$.

5. Choose a number $\overline{B} < B$ of independent replications and apply proposed test using one the available statistics (JB, SW or SW_0). $\overline{B}$ must be chosen to satisfy $T^{-1}\overline{B} \to 0$.

6. Reject null hypothesis of asymptotic normality (strong identification) if $p \leq \alpha$ where α is the nominal I type error and p is the p–value of normality test.

Alternatively, in order to avoid loss of information regarding $B - \overline{B}$ replications, it is possible to split the bootstrap sample of estimator in N non–overlapping groups of length $\overline{B} = int(B/N)$ and apply N, stochastically independent, normality tests. Thus, two decision rules may be applied:

- Bonferroni correction: select the value $\alpha_0 = \alpha N^{-1}$, and then reject null hypothesis if:

$$min_{i=1,\dots,N}(p_i) < \alpha_0.$$

- Sequential tests: select the overall I type error probability as $\alpha = 1 - (1 - \alpha_0)^N$, and again reject asymptotic normality of bootstrap estimator if $min_{i=1,\dots,N}(p_i) < \alpha_0$, where α_0 could be obtained by the following expression:

$$\alpha_0 = 1 - (1 - \alpha)^{1/N}.$$

For example, if the nominal level is $\alpha = 0.05$, the threshold for decision rule may be $\alpha_0 = 0.05/N$ or eventually $\alpha_0 = 1-(0.95)^{1/N}$. Considering a fixed number of group N,

[4] We point out that exogenous covariates could be removed before estimation, as explained in Section 2.1.3

the α_0 obtained using sequential tests is always greater than the Bonferroni correction version; moreover α_0 obtained through Bonferroni and that computed using (4.2.2) are very close when the number of groups N is large.

4.2.3 Standard normality tests

In order to extend the proposed approach, this section introduces some new statistics useful to verify the asymptotic (standard) normality of bootstrapped standardized TSLS estimators. These statistics are based on the moment conditions implied by standard normality; the main idea is to check if the moments of standardized bootstrapped estimators $\tilde{\beta}_T^*$ match those of standard normal, in order to detect if the non-studentized statistic $\tilde{\beta}_T$ is close or far from the $N(0,1)$. As pointed out in Chapter 2, non–normality of the standardized estimators appears weak instruments, often dramatically combined with the lack of moments, occurring in empirically–relevant cases of $k = 1$. However, even if the distribution of bootstrap estimator is normal, inferenec in TSLS may be affected by levels of endogeneity and degree of overidentification. Thus, if the null hypothesis of standard normality is safely rejected, this could be interpreted as a signal of weak instruments and, in general, misspecification regarding failures in the basic assumptions. All the proposed test statistics share a completely known χ^2 asymptotic distribution with known degrees of freedom from 1 to 4.

To introduce the methodology, let X a standard normal variable $X \sim N(0,1)$. Then, the main ingredients for test statistics are the following: firstly, the random vector $\mathbf{w}' = (X, X^2, X^3, X^4)$, its associated expectations $E(\mathbf{w})' = (0,1,0,3)$ and a completely known covariance matrix Ω equal to:

$$\Omega = \begin{pmatrix} 1 & 0 & 3 & 0 \\ 0 & 2 & 0 & 12 \\ 3 & 0 & 15 & 0 \\ 0 & 12 & 0 & 96 \end{pmatrix},$$

where the elements on principal diagonal are $V(X^j)$, where $j = 1, \ldots, 4$. In addition, $Cov(X^j, X^{j+1}) = 0 \ \forall j$ and

$$Cov(X, X^3) = 3; \quad Cov(X^2, X^4) = 12.$$

All proposed tests are built starting on the following statistic:

$$m_j = T^{-1/2}\left(\sum_{t=1}^{T} x_t^i - E(X^j)\right), \quad \text{where} \quad j = 1, \ldots, 4, \tag{4.7}$$

where $E(X^j)$ was previously defined and $D_T = (x_1, \ldots, x_t, \ldots, x_T)$ are the data[5]. Given the vector containing all ordered m_j, denoted with A_{1234}:

$$A_{1234} = \begin{pmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \end{pmatrix},$$

a comprehensive large samples test, involving the first four moments of the standard normal, could be based on the following expression:

$$M_{1234} = A'_{1234}\Omega^{-1}A_{1234} \xrightarrow{d} \chi^2_4. \tag{4.8}$$

where Ω is invertible by definition. Following expression in (4.8), it is possible to construct test statistics based on three moments of the standard normal using all the possible subvectors 3×1 of A_{1234}, denoted by $A_{jj'j''}$, and submatrices 3×3 from Ω, indicated with $\Omega_{jj'j''}$ for $j \neq j' \neq j''$. The generic statistic for test standard normality, using three moment conditions, is:

$$M_{jj'j''} = A'_{jj'j''}\Omega^{-1}_{jj'j''}A_{jj'j''} \xrightarrow{d} \chi^2_3, \tag{4.9}$$

where $j = \{1, 2\}$; $j' = \{2, 3\}$ and $j' > j$, $j'' = \{3, 4\}$ and $j'' > j'$. Four test statistics, obtainable from expression (4.9), are: $\{M_{123}, M_{124}, M_{134}, M_{234}\}$, where the subscripts denote statistics in the subvector $A_{jj'j''}$. Subsequentely, tests based on two moment conditions could be also implemented, considering subvectors 2×1 of A_{1234}, denoted by $A_{jj'}$, and submatrices 2×2 from Ω, i.e. $\Omega_{jj'}$. The generic test statistic $M_{jj'}$ is:

$$M_{jj'} = A'_{jj'}\Omega^{-1}_{jj'}A_{jj'} \xrightarrow{d} \chi^2_2, \tag{4.10}$$

where $j = \{1, 2, 3\}$, $j' = 2\,\{1, 2, 3\}$ and $j' > j$. Six test statistics could be obtained

[5] In this case the generic x_t is not an observation from endogenous regressor

from the (4.10) and are: $\{M_{12}, M_{13}, M_{14}, M_{23}, M_{24}, M_{34}\}$. Finally, it is possible to construct test statistics based on the single moments, simply using the (4.7):

$$M_j = \frac{m_j^2}{\omega_{jj}} \xrightarrow{d} \chi_1^2, \tag{4.11}$$

where $j = \{1, 2, 3, 4\}$ and ω_{jj} is the j-diagonal element of Ω, denoting the single $V(X^j)$; the four test statistics are: M_1, M_2, M_3, M_4.

Performance of the proposed tests is presented in Section 4.6 (Appendix) and their empirical size, power and asymptotic critical values are investigated through simulations. In order to evaluate empirical size and critical values of proposed tests, we generate 100000 samples and data x_i sampled from $N(0, 1)$, denoted as DGP_s, with different sizes $T = \{20, 50, 100, 200, 500, 1000, 5000, 10000\}$. Table 4.3 contains rejection frequencies regarding the 17 test statistics previously introduced. All tests present correct size in large samples, while M_{1234}, M_{234} seems to be slightly oversized when $T \leq 200$. The p–values are computed as $P(M \geq \chi^2_{df,1-\alpha})$, where the symbol "M" denotes one of the new proposed statistics, α is the selected nominal level equal to 0.05 and *df* is the number of degrees of freedom. Table 4.4 shows the 95% quantiles of these statistics, based on 1000000 replications, compared with the asymptotic critical values (χ_j^2) where $j = 1, \ldots, 4$. The difference between those values could be moderate in small samples and if number of moments involved in the test increases. Empirical power of proposed tests is investigated through two different data generating processes: DGP_{p1}, where $x_i \sim N(0.5, 1.2^2)$ and $DGP_{p2} : x_i \sim t(2)$. Main results are shown in Table 4.5 and 4.6. Under the DGP_{p1}, rejection frequencies are very close to 1 for all considered test, especially when the sample size is not too small ($T = 100$). When $x_t \sim t(2)$, rejections are very close to one for all tests (exept for M_1) even if $T = 50$.

Standard normality tests may be directly applied on $\overline{B} < B$ replications of the standardized IV/TSLS bootstrap estimates: $\tilde{\beta}^*_{T1}, \ldots, \tilde{\beta}^*_{T\overline{B}}$, when $\overline{B}T^{-1} \approx 0$ at the end of procedure described in the previous section. Then we could rewrite the test statistics in 4.7 in the following way:

$$m_j(\tilde{\beta}^*_T) = \overline{B}^{-1/2} \left(\sum_{b=1}^{\overline{B}} \tilde{\beta}^{*j}_{T,b} - E(X^j) \right), \quad \text{where} \quad j = 1, \ldots, 4, \tag{4.12}$$

where $\tilde{\beta}^*_T$ is the standardized bootstrapped IV,TSLS or $\kappa-$class estimator. P–values of the bootstrap–based test could be computed using the same procedures introduced in Section 4.2.2. To summarize, standard normality tests are computationally straightforward and do not require estimation of any covariance matrix, since Ω is completely known and does not depend on any nuisance parameter. Finally we remark that these tests are closely related to those proposed by Bontemps and Meddahi (2005), based on the Stein equations, obtained from the Hermite polynomials of the standard normal.

4.3 Monte Carlo simulation

In this section we present a small–scale Monte Carlo simulation, replicating $M = 1000$ dataset, in order to show the performances of proposed bootstrap–based tests with fixed and resampled instruments. Simulation design is the same proposed in Section 2.3.2, where instruments are drawn from multi–normal distribution $N_k(0, I_k)$ and the disturbances come from a multivariate normal $N_2(0, \mathbf{\Sigma})$. Again, different levels of positive endogeneity are considered, $\rho = \{0.25, 0.5, 0.75, 0.9\}$. The number of bootstrap replications is equal to $B = 199$ or $B = 399$, where tentative values for the subsamples $\overline{B}$ are $\overline{B} = (30, 40, 50)$ applying JB, SW and SW_0, while for standard normality tests we use $\overline{B} = (100, 120)$.

Firstly, we briefly analize diagnostic measures introduced in Section 4.1 using bootstrapped IV, TSLS and LIML. To our knowledge this is the first time that KS^* and MSE^* are applied to other $\kappa-$class estimators, and in general to the overidentified cases. Furthermore, we evaluate performance of new bootstrap–based test using conventional statistics (JB, SW), showing graphically their poor performances when the number of instruments increases. We also apply Shapiro–Wilk statistic with known mean (W_0) previously illustrated. The test is used also in the LIML case, especially because it shares the no–moment problem with the simple IV estimator. This means that under irrelevant and very weak instruments, proposed bootstrap normality test may be very powerful. Successively, we implement our proposed standard normality statistic in the overidentified cases. In particular we analize performance of the following statistics: $M_{12}, M_{13}, M_{123}, M_{124}, M_{134}, M_{1234}$. We use bold and red color to emphasize worst (power) and best (size) performance in our tables.

4.3.1 Diagnostics

Figure 4.3 shows the results of a small–scale simulation obtained by 1000 samples and then applying $B = 399$ bootstrap replications to compute KS^* and the mean of KS^* using $KS_1^*, \ldots, KS_M^*$. The graphic refers to the case of $k = 3$ instruments. The average values of KS^* are higher with weak identification and very high endogeneity, decreasing to 0.05 (tentative threshold) when instruments are collectively strong ($\mu^2/k \geq 40$). These results confirm the fact that $F > 10$, i.e. $\mu^2/k = 9$ could be a low benchmark, especially in case of high endogeneity. Table 4.7 shows mean and median of KS^* and median of MSE^* under weak instruments asymptotics, considering $\mu^2/k \in (0, 60)$ for both $k = 1$ (IV) and $k = 5$. We notice that median of bootstrapped MSE reaches huge values if $\mu^2/k \leq 10$, confirming the empirical threshold. Therefore, mean and median of bootstrapped KS increase, at the same level of identification, with high endogeneity.

Table 4.8 contains results regarding bootstrap mean square error and bootstrapped KS statistic for $\kappa-$class estimators LIML and Fuller, obtained through residual bootstrap with fixed instruments, under $k = 5$ and $T = 1000$. We show that, under irrelevant and weak instruments, the LIML may present huge values for the median of bootstrapped MSE. Fuller estimator performs better than LIML only in irrelevant and weak instrument cases, i.e. $\mu^2 \leq 5$, and also its bootstrapped mean square error decreases as the level of endogeneity increases, for all degrees of identification. Furthermore, KS^* rapidly decreases to 0.05 for all considered identification values. In general, under strong instruments, LIML and Fuller perform similarly in terms of MSE^* and KS^*. As we mentioned, these results confirm the difficult to identify a threshold in order to discriminate between strong instruments, although large values of MSE^* in IV and LIML estimator suggest the presence of irrelevant or very weak instruments.

4.3.2 Bootstrap–based tests

In Table 4.9 we present the results regarding normality tests in the just identified case, using the first $\overline{B}$ bootstrap replication of $\tilde{\beta}_T^{IV*}$, where $\overline{B} = 30, 40$ if $T = 100$ and $\overline{B} = 40, 50$ when $T = 500$. We consider only four identification cases; the tests present high rejection frequencies when instruments are weak or irrelevant, decreasing with the strength of instruments. However, rejection frequencies increase across endogeneity for

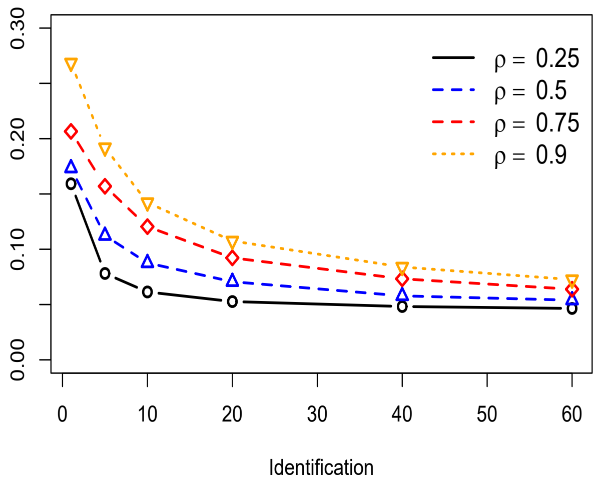
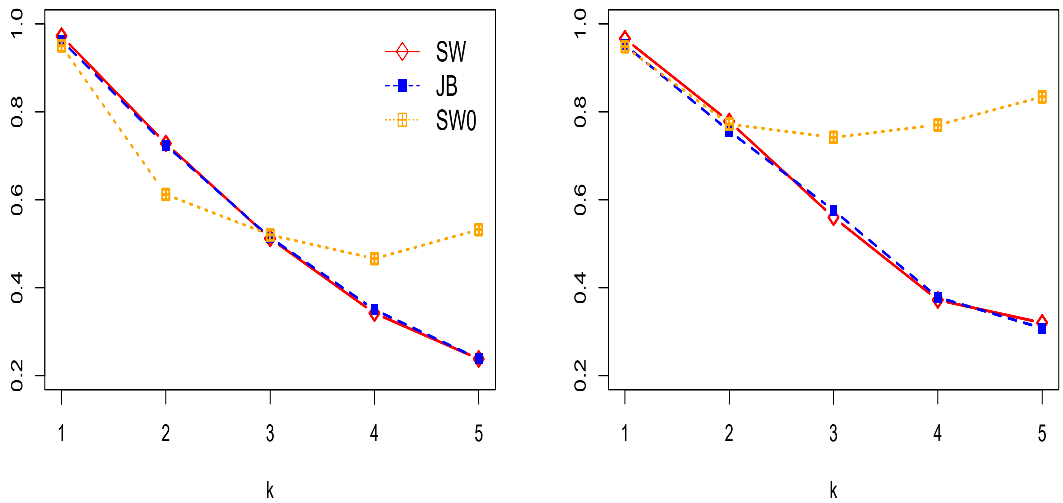


Figure 4.3: Mean of KS^* among different identification levels and degrees of endogeneity.

the same identification level, as required in IV framework. Table 4.11 contains main results of proposed tests when all the B replications are used, following the procedure described in Section 4.2.2. In this case, the empirical power increases under irrelevant and weak instruments, and the SW_0 test performs better than others in terms of empirical size, especially when instrument is very strong and endogeneity is moderately low e.g. $\rho = 0.25$ and $\mu^2 = 100$. In the overidentified models, i.e. when the number of instruments increases, empirical power of the bootstrap–based tests may be very low, especially under moderately small sample size ($T = 100$). To visualize this issue, in Figure 4.4 we plot the rejection frequencies of three bootstrap–based normality tests under irrelevant instruments, when $k = 5$. Conventional statistics decrease rapidly with the number of overidentification, while SW_0 statistic has power, under irrelevant instruments, especially when the endogeneity is high, as viewed in the left panel of the Figure. Nonetheless, in Table 4.10 the results of normality tests for TSLS and 3 instruments seem to remark the behaviour of IV case, in large sample size $T = 500$.

Considering $\kappa-$class estimators, Table 4.12 shows the performance of new bootstrap tests regarding LIML. We note that it performs very well in some cases, especially applying the procedure described in Section 4.2.2, and using the test statistic W_0 or the conventional W when endogeneity is not too high. In fact, empirical power is



(a) $\rho = 0.5$ (b) $\rho = 0.9$

Figure 4.4: Performance of three bootstrap–based normality tests among number of instrumental variables for two endogeneity levels, with $T = 100$ and $\overline{B} = 30$

equal to one under irrelevant instruments (lower panel of the Table), reaching often the nominal level when instruments are strong or very strong. To summarize, LIML estimator, presenting the moment problem for all possible k, seems useful to detect weak instruments because combination of no–moments and weak identification generates a detected non–normality.

Finally, we apply six of our proposed normality tests, in the overidentified case, inspired by evidences in Figure 4.4. Main results are illustrated in Table 4.13 under three possible scenarios, i.e. irrelevant, weak and strong instruments, corresponding to $\mu^2/k = 0, 1, 50$. The number of observations is $T = 500$, while proposed $\overline{B} = 100, 120$, due to asymptotically behaviour of tests shown in Table 4.3. We conclude that these tests tend to present high power under the combination of weak instruments ($\mu^2/k = 5$) and high endogeneity; in particular, when $\rho = 0.9$, the tests are more powerful under weak than irrelevant instruments. In general M_{12} and M_{123} seem to perform well among proposed statistics, presenting rejection frequencies closer to nominal level under the combination of strong instruments and low endogeneity. However, an optimal combination of $\overline{B}$ and T remains an issue for that we leave for further research.

4.4 Empirical Applications

In this section we give two example of discussed detection tools and new bootstrap tests using two different empirical datasets. The former represents an empirically–relevant case of one regressor–one instrument, while the second model presents four instruments and a set of control variables, helpful to improve inference in linear regression models. We apply Pair Bootstrap and three types of residual bootstrap: fixed instruments, resampled instruments and Residual Efficient bootstrap introduced by Davidson and MacKinnon (2010). The main idea is to see if the combination of a specific bootstrap method and estimator may help to detect weak identification.

Just identified case: Colonials Origins

The seminal paper of Acemoglu et al. (2001) is an example of instrumental variable estimation in Political Economics. The base sample size consists of $N = 64$ countries ex–European colonies; the outcome variable of interest is the logarithm of income per capita in 1995 (on the purchasing power parity basis), while the deemed endogenous regressor is an averaged index of risk protection against government appropriation of assets between 1985 and 1995. Authors suggest to use, as instrumental variable, the logarithm of mortality referring to European settlers during the colonization period. The just-identified model takes the following specification:

$$\begin{aligned} y_i &= \alpha + \beta R_i + u_i \\ x_i &= \tau + \pi m_i + v_i \end{aligned}$$

where y_i is the logarithm of DGP, R_i is the Risk index and m_i is the logarithm of mortality previoulsy discussed. Thus, we commonly estimate β using instrumental variable estimator defined in the (2.12). First stage results, $F = 22.4$ and $R_f^2 = 0.27$ suggest that instruments are not weak, especially because F value exceeds both the empirical threshold of ten and the critical values of Stock and Yogo (2005). The estimates of β is equal to 0.945 where estimated standard error is equal to 0.157.

In Table 4.1 there are two diagnostic tools and proposed bootstrap based tests through the four considered bootstrap methods. Pair bootstrap exhibits the larger MSE^*, confirming its inefficiency among other methods, while residual bootstrap with fixed instrument presents the lower value of KS^*. Despite these almost contradictory results, normality tests based on Shapiro–Wilk and Jarque–Bera statistics, based on

$\overline{B} = int(T/3)$, confirm normality of the bootstrap distribution giving another evidence of strong identification, previously checked by the first stage F.

Returns to schooling: Wage Equation

In order to apply bootstrap methods in a general framework consisting in more than one instrument and control variables, we consider the example of Card (2001), regarding the returns to education, i.e. quantifying the effect of an additional year of schooling on individual wage. Data comes from National Longitudinal Survey of Young Men (LSYM) between 1966-1981 including 3010 individuals; the outcome variable of interest is the logarithm of wage in 1976 while the endogenous regressor contains years of education even in 1976. Following wage equation is specified:

$$\begin{aligned}
\mathbf{y} &= \alpha_1 + \beta\mathbf{x} + \boldsymbol{\gamma}_1\mathbf{W}' + \mathbf{u} \\
\mathbf{x} &= \alpha_z + \mathbf{Z}\boldsymbol{\pi} + \boldsymbol{\gamma}_2\mathbf{W}' + \mathbf{v},
\end{aligned}$$

where $\mathbf{y}$ = logwage, $\mathbf{x}$ = education and β is the structural coefficient. We instrumented education with four dummy variables ($k = 4$) regarding college proximity, following the specification of Davidson and MacKinnon (2010): $\mathbf{Z} = (nearc2, nearc2 : nearc4, nearc4a, nearc4b)'$, respectively equal to 1 if there is a 2 year college, either a 2 and 4 year college, a four year pubblic college, and four year private college in the local area of each individual. Furthermore the matrix $\mathbf{W}$ contains $l = 5$ control variables including age and squared age, and three other dummy variables; we use the same controls of Davidson and MacKinnon (2010). The estimates of β is equal to 0.115 and first stage F is equal to 4.98; this suggests that instruments may be collectively weak.

In order to apply residual bootstrap with resampled instruments, we remove the effect of control variables and then estimate the overidentified model. Thus, we apply both TSLS and LIML estimators, since they are not equivalent when there are more instruments. In Table 4.2 we again present the diagnostic tools and p–values for bootstrap–based normality tests, computing using $\overline{B} = 50$ replications. As we can see, the results regarding TSLS are contradictory, since applying one method (Residual bootstrap with fixed instruments) the null hypothesis of strong instruments is not safely rejected. Moreover, using the LIML estimators all of bootstrap based tests reject the null hypothesis. This result confirms our findings in simulation study and in general the usefulness of LIML estimator that could be highly non–normal and

Method	KS^*	MSE^*	p–value(SW)	p–value(JB)
Pair	0.094	**5.614**	0.098	0.288
Residual (fix.)	0.051	0.034	0.202	0.470
Residual (res.)	0.093	0.447	0.725	0.877
Res Efficient	0.089	0.047	0.566	0.628

Table 4.1: Colonial Origins, where $B = 9999$ and $\overline{B} = T/3$

TSLS				
Method	KS^*	MSE^*	p–value(SW)	p–value(JB)
Pair	0.122	0.002	0.002	0.000
Residual (fix)	0.153	0.001	0.186	0.158
Residual (res.)	0.119	0.002	0.050	0.043
Res. Efficient	**0.519**	0.046	0.001	0.000

LIML				
Method	KS^*	MSE^*	p–value(SW)	p–value(JB)
Pair	0.106	0.002	0.000	0.000
Residual (fix)	0.065	0.001	0.006	0.002
Residual (res)	0.115	0.002	0.000	0.000
Res. Efficient	**0.328**	0.046	0.000	0.000

Table 4.2: Returns to Schooling, where $B = 9999$ and $\overline{B} = 50$

overdispersed if $\mu^2/k < 5$ (recalling that estimated first stage F is equal to 4.98). To summarize, in overidentified models we suggest that combination of new bootstrap based tests equipped with LIML estimator could be a useful screen in order to avoid weak identification.

4.5 Concluding Remarks

In this Chapter we exploit bootstrap methods to diagnose and test weak instruments in IV/TSLS inference; the usage of bootstrap (failures) to detect model misspecification represents a novel strand in the current literature. Firstly, we introduce graphical evaluation of (bootstrapped) distributions and other descriptive tools to detect weak identification applying bootstrap methods. Furthermore, we develop new bootstrap–based tests useful to (empirically) verify the relevance of instruments by testing asymptotic normality directly on the bootstrap replications of the estimator. We firstly apply well–known test statistics as Jarque Bera and Shapiro–Wilk, which seem to work well especially in the just identified case and when other no–moment estimators are applied in overidentified cases (e.g. LIML). Then, inspired by out simulation study, we propose test statistics for normality with fixed parameters, as the Shapiro–Wilk test with known mean and, finally, standard normality tests, for the null hypothesis that available data match the moments of $N(0,1)$. Simulation shows that some of these tests present a desirable behaviour in terms of size and power under weak instruments, being sensitive to some issue related to lack of moments, overidentification and, in particular, high endogeneity, which dramatically affects finite sample bias of IV/TSLS estimators. We also apply normality tests on two well–known datasets finding results consistent with the literature both in just–identified case and in more instruments situation.

These methods could be also quickly adapted to the case of multiple endogenous regressor. In fact, in order to evaluate relevance of instruments in each equation, a first strategy consists in testing (asymptotic) normality of each bootstrapped distribution of the estimators for the vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)'$. Alternatively, multivariate normality statistic of Doornik and Hansen (2008) may be applied to test the joint normality of $\hat{\boldsymbol{\beta}}_T^*$. We leave this topic for further resarch. Finally, the method can be straightforward applied of other $\kappa-$class estimators, as LIML or Fuller, considering more endogenous regressors.

4.6 Appendix

Moments of the Standard Normal

As indicated in section 4.2.3, we start considering $X \sim N(0,1)$, and then the random vector $\mathbf{w}' = (X, X^2, X^3, X^4)$. Expectation of X^j when $j = 1, \ldots, 4$ are given by:

$$E(X) = 0; \quad E(X^2) = 1; \quad E(X^3) = 0; \quad E(X^4) = 3$$

and the variances of elements included in $\mathbf{w}$ are:

$$\begin{aligned}
V(X) &= E(X^2) - [E(X)]^2 = & 1 - 0 = 1 \\
V(X^2) &= E(X^4) - [E(X^2)]^2 = & 3 - 1 = 2 \\
V(X^3) &= E(X^6) - [E(X^3)]^2 = & 15 - 0 = 15 \\
V(X^4) &= E(X^8) - [E(X^4)]^2 = & 105 - 9 = 96.
\end{aligned}$$

Then, the population covariances are:

$$Cov\left[X^i, X^{i+1}\right] = Cov\left[X^i, X^{i+d}\right] = 0.$$

are equal to zero if $d \in \{2, 4, \ldots\}$. In fact:

$$Cov\left[X, X^2\right] = Cov\left[X^2, X^3\right] = Cov\left[X^3, X^4\right] = Cov\left[X^1, X^4\right] = 0,$$

Furthermore there are two non–zero covariances:

$$\begin{aligned} Cov\left[X, X^3\right] &= E(X^4) - E(X)E(X^3) = 3 - 0 = 3, \\ Cov\left[X^2, X^4\right] &= E(X^6) - E(X^2)E(X^4) = 15 - 3 = 12. \end{aligned}$$

Thus, summarizing previous results for the random vector $\mathbf{w}$, it is possibile to rewrite:

$$\mathbf{w} = \begin{pmatrix} X \\ X^2 \\ X^3 \\ X^4 \end{pmatrix}, \quad E(\mathbf{w}) = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 3 \end{pmatrix}; \quad V(\mathbf{w}) = \Omega = \begin{pmatrix} 1 & 0 & 3 & 0 \\ 0 & 2 & 0 & 12 \\ 3 & 0 & 15 & 0 \\ 0 & 12 & 0 & 96 \end{pmatrix}.$$

Standard Normality Tests

We assume that data, indicated with $D_T = (x_1 \ldots, x_t, \ldots, x_T)'$ normally distributed under the null hypothesis, i.e. $x_t \sim N(0, 1)$

Test based on a single sample moment

Given:

$$\begin{cases} m_1 = \frac{1}{\sqrt{T}}\left(\sum_t x_t - E(X)\right) \\ m_2 = \frac{1}{\sqrt{T}}\left(\sum_t x_t^2 - E(X^2)\right) \\ m_3 = \frac{1}{\sqrt{T}}\left(\sum_t x_t^3 - E(X^3)\right) \\ m_4 = \frac{1}{\sqrt{T}}\left(\sum_t x_t^4 - E(X^4)\right) \end{cases} = \begin{cases} m_1 = \frac{1}{\sqrt{T}}\left(\sum_{t=1}^T x_t - 0\right) \\ m_2 = \frac{1}{\sqrt{T}}\left(\sum_{t=1}^T x_t^2 - 1\right) \\ m_3 = \frac{1}{\sqrt{T}}\left(\sum_{t=1}^T x_t^3 - 0\right) \\ m_4 = \frac{1}{\sqrt{T}}\left(\sum_{t=1}^T x_t^4 - 3\right), \end{cases}$$

we introduce four statistics based on sample moments:

$$M_1 = \frac{m_1^2}{V(X)} = m_1^2 \xrightarrow{d} \chi_1^2 \tag{4.13}$$

$$M_2 = \frac{m_2^2}{V(X^2)} = \frac{m_2^2}{2} \xrightarrow{d} \chi_1^2 \tag{4.14}$$

$$M_3 = \frac{m_3^2}{V(X^3)} = \frac{m_3^2}{15} \xrightarrow{d} \chi_1^2 \tag{4.15}$$

$$M_4 = \frac{m_4^2}{V(X^4)} = \frac{m_4^2}{96} \xrightarrow{d} \chi_1^2 \tag{4.16}$$

A comprehensive test of the four sample moments

A test statistic involving all the first four moments of standard normal is constructed using the quantity A:

$$A = \begin{pmatrix} \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t - E(X)\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^2 - E(X^2)\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^3 - E(X^3)\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^4 - E(X^4)\right) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t - 0\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^2 - 1\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^3 - 0\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^4 - 3\right) \end{pmatrix} \xrightarrow{d} N_4(0, \Omega) \tag{4.17}$$

where the covariance matrix previously defined is:

$$\Omega = \begin{pmatrix} 1 & 0 & 3 & 0 \\ 0 & 2 & 0 & 12 \\ 3 & 0 & 15 & 0 \\ 0 & 12 & 0 & 96 \end{pmatrix},$$

and a comprehensive test statistic has the following asymptotic distribution:

$$M_{1234} = A'\Omega^{-1}A \xrightarrow{d} \chi_4^2$$

Three sample moments tests

The main ingredients for built test statistics are the following four matrices $A_{jj'j''}$, where $j < j' < j''$:

$$A_{123} = \begin{pmatrix} \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t - E(X)\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^2 - E(X^2)\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^3 - E(X^3)\right) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t - 0\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^2 - 1\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^3 - 0\right) \end{pmatrix} \xrightarrow{d} N_3\left(0, \Omega_{123}\right) \quad (4.18)$$

$$A_{124} = \begin{pmatrix} \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t - E(X)\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^2 - E(X^2)\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^4 - E(X^4)\right) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t - 0\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^2 - 1\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^4 - 3\right) \end{pmatrix} \xrightarrow{d} N_3\left(0, \Omega_{124}\right) \quad (4.19)$$

$$A_{134} = \begin{pmatrix} \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t - E(X)\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^3 - E(X^3)\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^4 - E(X^4)\right) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t - 0\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^3 - 0\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^4 - 3\right) \end{pmatrix} \xrightarrow{d} N_3\left(0, \Omega_{134}\right) \quad (4.20)$$

$$A_{234} = \begin{pmatrix} \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^2 - E(X^2)\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^3 - E(X^3)\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^4 - E(X^4)\right) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^2 - 1\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^3 - 0\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^4 - 3\right) \end{pmatrix} \xrightarrow{d} N_3\left(0, \Omega_{234}\right) \quad (4.21)$$

where the four covariance matrices, indicated with $\Omega_{jj'j''}$ $j' > j$ and $j'' > j'$, are:

$$\Omega_{123} = \begin{pmatrix} 1 & 0 & 3 \\ 0 & 2 & 0 \\ 3 & 0 & 15 \end{pmatrix}; \quad \Omega_{124} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 12 \\ 0 & 12 & 96 \end{pmatrix};$$

$$\Omega_{134} = \begin{pmatrix} 1 & 3 & 0 \\ 3 & 15 & 0 \\ 0 & 0 & 96 \end{pmatrix}; \quad \Omega_{234} = \begin{pmatrix} 2 & 0 & 12 \\ 0 & 15 & 0 \\ 12 & 0 & 96 \end{pmatrix}.$$

Then the four test statistics, presenting an asymptotically χ^2_3 distribution, are the following:

$$\begin{aligned} M_{123} &= A'_{123}\Omega^{-1}_{123}A_{123} \xrightarrow{d} \chi^2_3 &(4.22)\\ M_{124} &= A'_{124}\Omega^{-1}_{124}A_{124} \xrightarrow{d} \chi^2_3 &(4.23)\\ M_{134} &= A'_{134}\Omega^{-1}_{134}A_{134} \xrightarrow{d} \chi^2_3 &(4.24)\\ M_{234} &= A'_{234}\Omega^{-1}_{234}A_{234} \xrightarrow{d} \chi^2_3 &(4.25)\end{aligned}$$

Tests based on two sample moments

The test statistics based on the vector $A_{jj'}$, where $j' > j$ are asymptotically normal, as follows:

$$A_{12} = \begin{pmatrix} \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t - E(X)\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^2 - E(X^2)\right) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t - 0\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^2 - 1\right) \end{pmatrix} \xrightarrow{d} N_2\left(0, \Omega_{12}\right)$$

$$A_{13} = \begin{pmatrix} \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t - E(X)\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^3 - E(X^3)\right) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t - 0\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^3 - 0\right) \end{pmatrix} \xrightarrow{d} N_2\left(0, \Omega_{13}\right)$$

$$A_{23} = \begin{pmatrix} \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^2 - E(X^2)\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^3 - E(X^3)\right) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^2 - 1\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^3 - 0\right) \end{pmatrix} \xrightarrow{d} N_2\left(0, \Omega_{23}\right)$$

$$A_{14} = \begin{pmatrix} \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t - E(X)\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^4 - E(X^4)\right) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t - 0\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^4 - 3\right) \end{pmatrix} \xrightarrow{d} N_2\left(0, \Omega_{14}\right)$$

$$A_{34} = \begin{pmatrix} \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^3 - E(X^3)\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^4 - E(X^4)\right) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^3 - 0\right) \\ \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} x_t^4 - 3\right) \end{pmatrix} \xrightarrow{d} N_2\left(0, \Omega_{34}\right) \quad (4.26)$$

where the six known covariance matrices, denoted as $\Omega_{jj'}$, are:

$$\begin{aligned}
\Omega_{12} &= \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}; \quad \Omega_{13} = \begin{pmatrix} 1 & 3 \\ 3 & 15 \end{pmatrix}; \quad \Omega_{14} = \begin{pmatrix} 1 & 0 \\ 0 & 96 \end{pmatrix}; \\
\Omega_{23} &= \begin{pmatrix} 2 & 0 \\ 0 & 15 \end{pmatrix}; \quad \Omega_{24} = \begin{pmatrix} 2 & 12 \\ 12 & 96 \end{pmatrix}; \quad \Omega_{34} = \begin{pmatrix} 15 & 0 \\ 0 & 96 \end{pmatrix}.
\end{aligned}$$

Finally, given the quantity $A_{jj'}$ and $\Omega_{jj'}$, where $j' > j$, the six test statistics (based on two moments implied by standard normal) are asymptotically χ^2 with 2 degrees of freedom, as follows:

$$M_{12} = A'_{12}\Omega_{12}^{-1}A_{12} \xrightarrow{d} \chi_2^2 \tag{4.27}$$

$$M_{13} = A'_{13}\Omega_{13}^{-1}A_{13} \xrightarrow{d} \chi_2^2 \tag{4.28}$$

$$M_{14} = A'_{14}\Omega_{14}^{-1}A_{14} \xrightarrow{d} \chi_2^2 \tag{4.29}$$

$$M_{23} = A'_{23}\Omega_{23}^{-1}A_{23} \xrightarrow{d} \chi_2^2 \tag{4.30}$$

$$M_{24} = A'_{24}\Omega_{24}^{-1}A_{24} \xrightarrow{d} \chi_2^2 \tag{4.31}$$

$$M_{34} = A'_{34}\Omega_{34}^{-1}A_{34} \xrightarrow{d} \chi_2^2 \tag{4.32}$$

An R code to obtain bootstrap–based normality tests

Here we present a simple R function to compute bootstrap–based normality test through procedure described in Section 4.2.2. We include only SW and JB test applied on IV/TSLS estimator. Code could be straigthforward extended to other $\kappa-$class estimators.

```
boot_test<-function(B,b,data){
library(tseries) ## to compute JB test
k<-ncol(data)-2
t<-nrow(data)
iv<-ivreg(data[,1]~data[,2]|data[,3:(k+2)])
thetaB<-numeric()
for(j in 1:B){
indices<-sample(1:t,t,replace=TRUE)
## Residual bootstrap with resampled instruments
uvz<-cbind(residuals(iv),
residuals(lm(data[,2]~data[,3:(k+2)])),data[,3:(k+2)])
ub<-uvz[indices,1];vb<-uvz[indices,2];zb<-uvz[indices,3:(k+2)]
if (k==1) xb<-lm(data[,2]~data[,3:(k+2)])$coef[2:(k+1)]*zb+vb
else xb<-lm(data[,2]~data[,3:(k+2)])$coef[2:(k+1)]%*%zb+vb
yb<-iv$coef[2]*xb+ub
ivb<-ivreg(yb~xb|zb)
thetaB[j]<-ivb$coef[2]
}
## Standardized coefficient
thetaBst<-(thetaB-iv$coef[2])/summary(iv)$coefficients[2,2]
## Split in N groups
cfa<-split(thetaBst, sort(rank(thetaBst) %% round(B/b)))
p1s<-numeric();p2s<-numeric();
for(wi in 1:length(cfa)){
p1s[wi]<-shapiro.test(cfa[[wi]])$p.value;
p2s[wi]<-jarque.bera.test(cfa[[wi]])$p.value
}
## p.values with our procedure
```

```
p1<-min(p1s);p2<-min(p2s)
##threshold following Section 4.2.2
tr<-1-(0.95)^(1/round(B/b))
out<-c(p1<tr,p2<tr)
names(out)<-c("SW","JB")
return(out)}
```

	M_j				$M_{jj'}$						$M_{jj'j''}$				
T	M_1	M_2	M_3	M_4	M_{12}	M_{13}	M_{14}	M_{23}	M_{24}	M_{34}	M_{123}	M_{124}	M_{134}	M_{234}	M_{1234}
20	0.048	0.046	0.057	0.046	0.048	0.059	0.054	0.058	0.050	0.061	0.061	0.054	0.064	0.061	0.062
50	0.049	0.047	0.050	0.040	0.046	0.055	0.048	0.050	0.047	0.054	0.055	0.051	0.058	0.056	0.059
100	0.046	0.049	0.053	0.046	0.046	0.056	0.050	0.051	0.056	0.055	0.054	0.053	0.060	0.061	0.063
200	0.052	0.045	0.053	0.045	0.050	0.056	0.051	0.048	0.051	0.051	0.052	0.055	0.055	0.056	0.062
500	0.051	0.049	0.052	0.046	0.050	0.050	0.050	0.050	0.048	0.050	0.049	0.052	0.051	0.052	0.053
1000	0.047	0.051	0.052	0.050	0.051	0.055	0.049	0.053	0.053	0.051	0.050	0.052	0.051	0.053	0.054
5000	0.049	0.054	0.052	0.053	0.051	0.049	0.050	0.051	0.055	0.052	0.050	0.051	0.051	0.054	0.052
10000	0.048	0.054	0.049	0.054	0.050	0.049	0.055	0.051	0.055	0.056	0.054	0.054	0.055	0.052	0.055

Table 4.3: Rejection Frequency of standard normality tests. $DGP_s : x_i \sim N(0,1)$

	M_j				$M_{jj'}$						$M_{jj'j''}$				
T	M_1	M_2	M_3	M_4	M_{12}	M_{13}	M_{14}	M_{23}	M_{24}	M_{34}	M_{123}	M_{124}	M_{134}	M_{234}	M_{1234}
20	3.805	3.662	4.154	3.581	5.967	6.529	6.357	6.685	6.111	7.187	8.762	8.249	9.255	9.119	11.009
50	3.818	3.755	4.032	3.528	5.900	6.401	6.203	6.248	6.222	6.827	8.409	8.272	8.893	8.874	10.815
100	3.793	3.805	3.906	3.511	5.895	6.184	6.016	6.036	6.194	6.341	8.044	8.117	8.444	8.584	10.441
200	3.844	3.843	3.901	3.663	5.994	6.191	6.010	6.047	6.173	6.106	8.025	8.115	8.234	8.356	10.213
500	3.855	3.846	3.883	3.761	6.008	6.051	5.986	5.995	6.028	6.010	7.872	7.902	7.947	8.013	9.793
1000	3.842	3.877	3.825	3.815	6	6.014	6.042	5.995	6.054	6.005	7.826	7.935	7.907	7.956	9.690
5000	3.857	3.860	3.839	3.866	6.027	6.002	6.051	6.018	6.039	6.032	7.860	7.857	7.883	7.856	9.533
10000	3.854	3.884	3.842	3.857	6.008	6.001	6.007	5.995	6.040	5.991	7.861	7.860	7.856	7.861	9.547
	$\chi^2_{1,0.95} \cong 3.84$				$\chi^2_{2,0.95} \cong 5.99$						$\chi^2_{3,0.95} \cong 7.81$				$\chi^2_{4,0.95} \cong 9.48$

Table 4.4: 95%-quantile of the proposed test statistics. The $DGP_{s1} : x_i \sim N(0,1)$.

	M_j				$M_{jj'}$						$M_{jj'j''}$				
T	M_1	M_2	M_3	M_4	M_{12}	M_{13}	M_{14}	M_{23}	M_{24}	M_{34}	M_{123}	M_{124}	M_{134}	M_{234}	M_{1234}
20	0.599	0.508	0.594	0.471	0.672	0.607	0.654	0.642	0.490	0.605	0.662	0.650	0.631	0.621	0.641
50	0.908	0.808	0.871	0.750	0.948	0.899	0.938	0.918	0.780	0.896	0.938	0.936	0.922	0.905	0.928
100	0.995	0.969	0.981	0.941	0.999	0.991	0.998	0.995	0.957	0.991	0.997	0.998	0.996	0.993	0.997
200	1.000	1.000	1.000	0.998	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000
500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 4.5: Rejection frequencies of normality tests(empirical power). $DGP_{p1} : x_i \sim N(0.5, 1.2^2)$.

	M_j				$M_{jj'}$						$M_{jj'j''}$				Moments= 4
T	M_1	M_2	M_3	M_4	M_{12}	M_{13}	M_{14}	M_{23}	M_{24}	M_{34}	M_{123}	M_{124}	M_{134}	M_{234}	M_{1234}
20	0.363	0.850	0.780	0.880	0.840	0.795	0.875	0.872	0.873	0.885	0.872	0.871	0.882	0.882	0.879
50	0.414	0.988	0.914	0.994	0.986	0.933	0.992	0.991	0.992	0.993	0.990	0.992	0.993	0.993	0.993
100	0.437	1.000	0.957	1.000	1.000	0.969	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
200	0.462	1.000	0.980	1.000	1.000	0.987	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
500	0.495	1.000	0.990	1.000	1.000	0.994	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1000	0.515	1.000	0.996	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 4.6: Rejection frequencies of standard normality tests(empirical power). $DGP_{p2} : x_i \sim t(2)$.

		$k=1$			$k=5$		
ρ	μ^2/k	Me(MSE^*)	$\overline{KS}^*$	Me(KS^*)	Me(MSE^*)	$\overline{KS}^*$	Me(KS^*)
0.25	0	**496.393**	0.220	0.183	0.271	0.202	0.179
	1	**258.961**	0.173	0.112	0.132	0.143	0.118
	5	9.768	0.093	0.077	0.038	0.088	0.078
	10	0.252	0.065	0.060	0.019	0.076	0.072
	20	0.064	0.054	0.051	0.010	0.061	0.058
	40	0.028	0.048	0.046	0.005	0.054	0.051
	60	0.018	0.048	0.046	0.003	0.053	0.051
0.5	0	**378.260**	0.216	0.175	0.219	0.201	0.175
	1	**229.034**	0.177	0.114	0.115	0.163	0.148
	5	11.583	0.105	0.086	0.036	0.143	0.136
	10	0.359	0.076	0.069	0.019	0.117	0.114
	20	0.068	0.061	0.058	0.010	0.094	0.093
	40	0.028	0.053	0.052	0.005	0.075	0.075
	60	0.018	0.050	0.049	0.003	0.067	0.066
0.75	0	204.041	0.217	0.183	0.128	0.201	0.178
	1	194.010	0.186	0.129	0.085	0.223	0.207
	5	21.000	0.114	0.098	0.033	0.204	0.204
	10	0.476	0.091	0.087	0.018	0.167	0.166
	20	0.075	0.073	0.070	0.009	0.128	0.127
	40	0.030	0.060	0.059	0.005	0.102	0.102
	60	0.018	0.055	0.053	0.003	0.088	0.088
0.9	0	96.611	0.223	0.187	0.052	0.202	0.176
	1	179.791	0.212	0.154	0.070	0.294	0.287
	5	24.064	0.128	0.112	0.030	0.252	0.245
	10	0.621	0.102	0.100	0.018	0.201	0.200
	20	0.082	0.082	0.079	0.009	0.153	0.154
	40	0.030	0.066	0.064	0.005	0.116	0.116
	60	0.019	0.060	0.058	0.003	**0.100**	0.100

Table 4.7: MSE^* and KS^* for IV/TSLS estimators with $k = 1, 5$ with different strength of instruments and endogeneity levels.

		LIML			Fuller($\underline{c}=1$)		
ρ	μ^2/k	Me(MSE^*)	$\overline{KS}^*$	Me(KS^*)	Me(MSE^*)	$\overline{KS}^*$	Me(KS^*)
	0	**354.415**	0.172	0.172	0.655	0.145	0.134
	1	78.046	0.136	0.136	0.364	0.119	0.110
	5	0.180	0.086	0.086	0.074	0.081	0.075
0.25	10	0.030	0.064	0.064	0.028	0.064	0.060
	20	0.012	0.054	0.054	0.012	0.056	0.053
	40	0.005	0.051	0.051	0.005	0.053	0.050
	60	0.003	0.050	0.050	0.003	0.053	0.050
	0	296.829	0.172	0.172	0.477	0.142	0.133
	1	84.282	0.141	0.141	0.294	0.123	0.112
	5	0.109	0.082	0.082	0.065	0.078	0.074
0.5	10	0.029	0.064	0.064	0.026	0.065	0.061
	20	0.012	0.055	0.055	0.011	0.057	0.054
	40	0.005	0.052	0.052	0.005	0.054	0.052
	60	0.003	0.050	0.050	0.003	0.053	0.051
	0	153.072	0.166	0.166	0.279	0.143	0.133
	1	72.258	0.141	0.141	0.214	0.132	0.112
	5	0.089	0.084	0.084	0.054	0.074	0.071
0.75	10	0.026	0.065	0.065	0.024	0.065	0.062
	20	0.011	0.058	0.058	0.011	0.059	0.057
	40	0.005	0.054	0.054	0.005	0.056	0.053
	60	0.003	0.052	0.052	0.003	0.055	0.053
	0	82.135	0.179	0.179	0.132	0.148	0.142
	1	60.265	0.131	0.131	0.152	0.157	0.125
	5	0.068	0.080	0.080	0.047	0.079	0.077
0.9	10	0.027	0.067	0.067	0.022	0.069	0.066
	20	0.011	0.057	0.057	0.010	0.061	0.058
	40	0.005	0.054	0.054	0.005	0.057	0.055
	60	0.003	0.053	0.053	0.003	0.055	0.053

Table 4.8: MSE^* and KS^* for LIML and Fuller estimators with different strength of instruments and endogeneity levels, $k = 5$.

Resampled instrument

		$T = 100$					
		$\overline{B}_1 = 30$			$\overline{B}_2 = 40$		
ρ	μ^2	SW	JB	SW_0	SW	JB	SW_0
0.25	0	0.940	0.918	0.901	0.978	0.973	0.949
	1	0.902	0.886	0.851	0.957	0.954	0.917
	50	0.116	0.111	0.089	0.145	0.142	0.110
	100	0.098	0.066	0.063	0.090	0.080	0.075
0.25	0	0.942	0.917	0.909	0.973	0.965	0.952
	1	0.899	0.875	0.862	0.935	0.931	0.900
	50	0.174	0.140	0.119	0.214	0.179	0.137
	100	0.127	0.103	0.084	0.123	0.116	0.081
0.25	0	0.943	0.927	0.895	0.969	0.969	0.949
	1	0.951	0.925	0.918	0.964	0.961	0.947
	50	0.325	0.256	0.204	0.458	0.372	0.275
	100	0.198	0.144	0.117	0.241	0.207	0.163
		$T = 500$					
		$\overline{B}_1 = 40$			$\overline{B}_2 = 50$		
ρ	μ^2	SW	JB	SW_0	SW	JB	SW_0
0.25	0	0.982	0.973	0.965	0.983	0.977	0.952
	1	0.931	0.926	0.892	0.947	0.940	0.911
	50	0.139	0.138	0.091	0.139	0.125	0.100
	100	0.070	0.063	0.075	0.100	0.080	0.074
0.5	0	0.978	0.971	0.961	0.975	0.963	0.958
	1	0.949	0.937	0.914	0.942	0.941	0.915
	50	0.194	0.167	0.123	0.231	0.209	0.133
	100	0.126	0.087	0.084	0.109	0.102	0.105
0.9	0	0.983	0.983	0.965	0.973	0.966	0.948
	1	0.968	0.962	0.951	0.972	0.960	0.950
	50	0.206	0.192	0.124	0.235	0.214	0.142
	100	0.256	0.196	0.136	0.265	0.202	0.165

Fixed Instrument

		$T = 100$					
		$\overline{B}_1 = 30$			$\overline{B}_2 = 40$		
ρ	μ^2	SW	JB	SW_0	SW	JB	SW_0
0.25	0	0.945	0.931	0.919	0.971	0.962	0.943
	1	0.912	0.904	0.862	0.945	0.938	0.909
	50	0.119	0.110	0.080	0.147	0.129	0.099
	100	0.085	0.065	0.066	0.082	0.067	0.061
0.5	0	0.946	0.928	0.920	0.978	0.964	0.957
	1	0.905	0.892	0.852	0.941	0.939	0.917
	50	0.171	0.146	0.110	0.213	0.182	0.112
	100	0.105	0.081	0.077	0.128	0.102	0.090
0.9	0	0.940	0.930	0.911	0.976	0.972	0.948
	1	0.934	0.918	0.912	0.978	0.971	0.955
	50	0.334	0.262	0.218	0.404	0.337	0.248
	100	0.192	0.138	0.119	0.220	0.163	0.137
		$T = 500$					
		$\overline{B}_1 = 40$			$\overline{B}_2 = 50$		
ρ	μ^2	SW	JB	SW_0	SW	JB	SW_0
0.25	0	0.978	0.968	0.951	0.980	0.979	0.950
	1	0.950	0.946	0.912	0.946	0.936	0.901
	50	0.122	0.115	0.082	0.119	0.114	0.072
	100	0.067	0.067	0.051	0.085	0.068	0.046
0.5	0	0.981	0.979	0.956	0.979	0.971	0.951
	1	0.947	0.947	0.920	0.943	0.941	0.906
	50	0.210	0.189	0.130	0.232	0.190	0.124
	100	0.119	0.089	0.081	0.124	0.104	0.071
0.9	0	0.973	0.966	0.966	0.975	0.973	0.955
	1	0.970	0.958	0.956	0.973	0.965	0.957
	50	0.215	0.187	0.126	0.210	0.178	0.120
	100	0.218	0.174	0.125	0.256	0.192	0.138

Table 4.9: Bootstrap–based normality tests on the IV estimator under two different sample sizes and $B = 199$.

Resampled instruments

		$\overline{B}_1 = 40$			$\overline{B}_2 = 50$		
ρ	μ^2	SW	JB	SW_0	SW	JB	SW_0
0.25	0	0.982	0.973	0.965	**0.989**	0.988	0.970
	1	0.931	0.926	0.892	0.966	0.964	0.939
	50	0.139	0.138	0.091	0.146	0.138	0.091
	100	0.070	0.063	0.075	0.086	0.073	0.057
0.5	0	0.978	0.971	0.961	0.986	0.985	0.973
	1	0.949	0.937	0.914	0.962	0.964	0.938
	50	0.194	0.167	0.123	0.259	0.231	0.139
	100	0.126	0.087	0.084	0.147	0.131	0.082
0.9	0	0.983	0.983	0.965	0.992	0.993	0.978
	1	0.968	0.962	0.951	0.990	0.985	0.972
	50	0.206	0.192	0.124	0.238	0.218	0.142
	100	0.256	0.196	0.136	0.286	0.227	0.149

Fixed Instruments

		$\overline{B}_1 = 40$			$\overline{B}_2 = 50$		
ρ	μ^2	SW	JB	SW_0	SW	JB	SW_0
0.25	0	0.978	0.968	0.951	**0.985**	0.984	0.978
	1	0.950	0.946	0.912	0.957	0.961	0.932
	50	0.122	0.115	0.082	0.137	0.136	0.085
	100	0.067	0.067	0.051	0.091	0.078	0.046
0.5	0	0.981	0.979	0.956	0.988	0.984	0.975
	1	0.947	0.947	0.920	0.967	0.962	0.939
	50	0.210	0.189	0.130	0.252	0.231	0.125
	100	0.119	0.089	0.081	0.130	0.107	0.061
0.9	0	0.973	0.966	0.966	0.995	0.989	0.982
	1	0.970	0.958	0.956	0.977	0.974	0.970
	50	0.215	0.187	0.126	0.244	0.211	0.146
	100	0.218	0.174	0.125	0.288	0.227	0.149

Table 4.10: Bootstrap–based normality tests on the TSLS ($k = 3$) estimator with $T = 500$

Resampled instrument

		$T = 100$					
		$\overline{B}_1 = 30$			$\overline{B}_2 = 40$		
ρ	μ^2	SW	JB	SW_0	SW	JB	SW_0
0.25	0	0.998	1.000	0.994	1.000	1.000	0.996
	1	0.988	0.992	0.970	0.992	0.994	0.978
	50	0.258	0.332	0.086	0.262	0.366	0.138
	100	0.134	0.222	0.082	0.180	0.268	0.078
0.5	0	1.000	1.000	0.990	1.000	1.000	1.000
	1	0.996	0.998	0.984	1.000	1.000	0.988
	50	0.366	0.466	0.180	0.376	0.444	0.190
	100	0.198	0.292	0.092	0.238	0.300	0.108
0.9	0	1.000	1.000	0.992	1.000	1.000	0.996
	1	0.996	1.000	0.988	0.994	0.994	0.990
	50	0.622	0.670	0.298	0.750	0.772	0.444
	100	0.396	0.450	0.168	0.472	0.506	0.204

		$T = 500$					
		$\overline{B}_1 = 40$			$\overline{B}_2 = 50$		
ρ	μ^2	SW	JB	SW_0	SW	JB	SW_0
0.25	0	0.979	0.974	0.952	0.991	0.989	0.979
	1	0.937	0.932	0.902	0.996	0.996	0.984
	50	0.105	0.099	0.081	0.146	0.156	0.097
	100	0.077	0.068	0.052	0.150	0.176	0.082
0.5	0	0.967	0.959	0.945	0.992	0.989	0.984
	1	0.939	0.942	0.910	0.992	0.994	0.976
	50	0.212	0.188	0.109	0.229	0.205	0.133
	100	0.125	0.111	0.069	0.234	0.304	0.080
0.9	0	0.981	0.976	0.965	0.987	0.982	0.981
	1	0.954	0.948	0.937	0.981	0.980	0.959
	50	0.211	0.179	0.154	0.482	0.387	0.305
	100	0.245	0.185	0.118	0.307	0.249	0.162

Fixed Instrument

		$T = 100$					
		$\overline{B}_1 = 30$			$\overline{B}_2 = 40$		
ρ	μ^2	SW	JB	SW_0	SW	JB	SW_0
0.25	0	1.000	1.000	0.998	0.998	1.000	0.998
	1	0.988	0.994	0.970	0.994	0.996	0.984
	50	0.198	0.274	0.080	0.240	0.310	0.098
	100	0.100	0.166	0.078	0.108	0.176	0.068
0.5	0	0.996	0.998	0.992	1.000	1.000	1.000
	1	0.996	0.996	0.978	0.990	0.994	0.986
	50	0.322	0.438	0.138	0.348	0.428	0.184
	100	0.168	0.244	0.102	0.210	0.274	0.106
0.9	0	0.998	1.000	0.998	1.000	1.000	0.996
	1	0.998	0.998	0.988	0.998	0.998	0.992
	50	0.676	0.714	0.304	0.724	0.730	0.394
	100	0.346	0.448	0.126	0.398	0.480	0.206

		$T = 500$					
		$\overline{B}_1 = 40$			$\overline{B}_2 = 50$		
ρ	μ^2	SW	JB	SW_0	SW	JB	SW_0
0.25	0	0.976	0.973	0.963	0.988	0.983	0.978
	1	0.950	0.941	0.910	0.996	0.998	0.988
	50	0.126	0.111	0.084	0.143	0.145	0.083
	100	0.087	0.070	0.073	0.138	0.190	0.048
0.5	0	0.971	0.962	0.955	0.987	0.987	0.980
	1	0.946	0.937	0.906	0.990	0.996	0.984
	50	0.221	0.189	0.114	0.238	0.201	0.133
	100	0.132	0.106	0.079	0.238	0.300	0.082
0.9	0	0.974	0.969	0.958	0.989	0.985	0.978
	1	0.966	0.965	0.951	0.979	0.975	0.960
	50	0.211	0.191	0.141	0.522	0.420	0.327
	100	0.261	0.208	0.131	0.294	0.244	0.130

Table 4.11: Bootstrap–based normality tests on the IV estimator with two different sample size using procedure described in the text, with $B = 199$.

Test based on $\overline{B} < B$

		$\overline{B}_1 = 40$			$\overline{B}_2 = 50$		
ρ	μ^2/k	SW	JB	SW_0	SW	JB	SW_0
0.25	0	0.963	0.958	0.920	0.970	0.969	0.937
	1	0.796	0.790	0.697	0.854	0.859	0.782
	50	0.054	0.047	0.056	0.061	0.051	0.040
	100	0.055	0.039	0.055	0.059	0.049	0.056
0.5	0	0.940	0.933	0.900	0.979	0.977	0.947
	1	0.821	0.807	0.723	0.859	0.863	0.785
	50	0.073	0.057	0.044	0.083	0.055	0.052
	100	0.057	0.044	0.053	0.062	0.038	0.059
0.9	0	0.945	0.940	0.900	0.977	0.977	0.953
	1	0.916	0.868	0.838	0.930	0.899	0.874
	50	0.126	0.101	0.093	0.121	0.101	0.071
	100	0.073	0.047	0.050	0.105	0.083	0.066

Procedure based on overall B=199

		$\overline{B}_1 = 40$			$\overline{B}_2 = 50$		
ρ	μ^2/k	SW	JB	SW_0	SW	JB	SW_0
0.25	0	1.000	1.000	0.996	1.000	1.000	0.990
	1	0.968	0.984	0.928	0.980	0.990	0.944
	50	0.072	0.124	0.054	0.080	0.108	0.058
	100	0.052	0.098	0.036	0.060	0.096	0.046
0.5	0	0.992	0.992	0.986	0.979	0.977	0.947
	1	0.970	0.984	0.928	0.984	0.986	0.932
	50	0.074	0.128	0.052	0.118	0.168	0.048
	100	0.096	0.132	0.052	0.078	0.090	0.050
0.9	0	0.945	0.940	0.900	0.977	0.977	0.953
	1	0.916	0.868	0.838	0.930	0.899	0.874
	50	0.126	0.101	0.093	0.121	0.101	0.071
	100	0.073	0.047	0.050	0.105	0.083	0.066

Table 4.12: Boostrap–based tests and LIML estimator with $T = 500$ and $k = 5$ instruments

$\overline{B} = 100$

ρ	Identification	M_{12}	M_{12}	M_{123}	M_{124}	M_{134}	M_{1234}
	irrelevant	0.787	0.689	0.808	0.855	0.723	0.853
0.25	weak	0.582	0.596	0.638	0.656	0.589	0.670
	strong	0.049	0.064	0.060	0.060	0.064	0.068
	irrelevant	0.792	0.683	0.814	0.844	0.712	0.840
0.5	weak	0.678	0.675	0.718	0.727	0.667	0.737
	strong	0.072	0.109	0.099	0.076	0.102	0.099
	irrelevant	0.771	0.686	0.809	0.847	0.713	0.850
0.9	weak	0.952	0.959	**0.968**	0.949	0.956	0.965
	strong	0.166	0.248	0.215	0.144	0.206	0.179

$\overline{B} = 120$

ρ	Identification	M_{12}	M_{12}	M_{123}	M_{124}	M_{134}	M_{1234}
	irrelevant	0.821	0.718	0.850	0.891	0.750	0.887
0.25	weak	0.611	0.609	0.661	0.723	0.615	0.724
	strong	0.065	0.073	0.063	0.063	0.065	0.071
	irrelevant	0.833	0.763	0.877	0.916	0.794	0.917
0.5	weak	0.701	0.707	0.742	0.781	0.710	0.789
	strong	0.093	0.129	0.111	0.090	0.109	0.100
	irrelevant	0.811	0.701	0.837	0.892	0.735	0.891
0.9	weak	0.967	0.971	0.974	0.971	0.969	**0.975**
	strong	0.193	0.284	0.247	0.173	0.247	0.216

Table 4.13: Boostrap–based standard normality tests for TSLS estimator with $T = 500$ and $k = 5$ instruments under three identification scenarios and three levels of endogeneity

Chapter 5

Conclusions and further research

5.1 Summary and conclusions

In this PhD Thesis we analyze instrumental variable estimators and their bootstrap counterpart under poorly relevant instruments. In Chapter 2, we discuss how the weakness of excluded instruments affects both finite sample and asymptotic properties of IV/TSLS, and also of other $\kappa-$class estimators. We consider several weak instruments scenarios, reporting the performance of (non–robust) estimators and tests under standard and non–standard conditions, considering models with one endogenous regressor equipped with one or more instruments. Our simulations confirms that IV and TSLS may perform very poorly in terms of coverage rates and may generally be very far from their limiting distribution, even if the sample size is moderately large, mainly due a sensitivity to the level of endogeneity. LIML and Fuller estimators, nested in the so–called $\kappa-$class, perform better than IV/TSLS in terms of median point estimates and coverage rates; we also suggest their use in case of nearly-weak instruments and high degree of endogeneity, i.e. $5 < F < 10$ and $\rho > 0.5$. Nevertheless, both IV and LIML suffer from the no–moment problem and this is reflected in their huge variability, occurring especially when instruments are very weak or practically irrelevant. In this context, we propose KS statistics, suggested by Zhan (2017) in a different framework, as descriptive indicator of normal approximation, in order to compare the sensitivity of $\kappa-$class estimators with respect to misspecification. Moreover, when instruments are weakly endogenous, i.e. poorly correlated with structural disturbances, rejection frequencies of t/Wald test, associated to all $\kappa-$class estimators, may be too high even if instruments are not deemed weak. This will be also investigated in other test statistics mainly used in IV framework.

Bootstrap methods may often improve inference in finite samples. For this reason, different methods summarized in section 2.4 could be applied in linear IV setting. However, we prove that bootstrap counterpart of IV and TSLS estimator under weak instrument asymptotics converges to a non-normal distribution presenting random component, conditionally on the original data D_T. In the empirically relevant just–identified case, when a single instrument is totally irrelevant, this randomness is exacerbated due to the no–moment problem of IV estimator. Similar results may be found applying different bootstrap methods, with some exceptions, as pointed out in section 3.2.5. Our Monte Carlo exercises suggest that randomness can also be found in the bootstrap distribution of t–statistic associated with IV estimator and in bootstrapped $\kappa-$class estimators.

In Section 3.3 we present some cases of malfunctions in bootstrap inference regarding bias–correction, confidence intervals, and hypothesis testing under weak instruments. We also highlight cases in which bootstrap may perform better, in terms of empirical power or coverage rates, with respect to conventional asymptotics, with the exception of irrelevant or nearly–irrelevant instruments.

Furthemore, in Chapter 4 we introduce bootstrap–based normality tests in order to test the null hypothesis of strong instruments, using a limited number of bootstrap replications. Our simulation study in Section 4.3 shows that normality tests perform satistfyingly in terms of power through IV and LIML estimators, and procedure involving all the bootstrap replication, introduecd in Section 4.2.2, may also improve the power of test. In particular, the LIML shown desirable features in this context because its no–moment problem exacerbates non–normality of bootstrap replications under very weak and irrelevant instruments, producing better results in terms both of empirical size under strong instruments, and empirical power under irrelevant and weak instruments. These findings suggest to extend out test in the GMM framework on the Generalized Empirical Likelihood (GEL) estimator, reducing to LIML in the linear context. Finally, adding more instruments, rejection frequencies of conventional normality tests tend to decrease using TSLS estimator. To overcome this problem, we propose to test standard normality of bootstrap using different test statistics. These tests improve the empirical power when weak instruments are combined with high endogeneity, that often affects finite sample bias of TSLS.

5.2 Suggestions for further works

In order to develop the evaluation of Chapter 2, we will consider $\kappa-$class estimators in the case of multiple endogenous regressor under weak instruments, where the vector $\mathbf{\Pi}$ will be modeled in terms of population Cragg–Donald statistic (Cragg and Donald, 1993) or partial μ^2, using framework described in Angrist and Pischke (2008). Nowadays a question of practical interest regards the behaviour of partial F statistic, proposed by Sanderson and Windmeijer (2016), under non–standard conditions, i.e. heteroskedastic/autocorrelated disturbances and weakly endogenous instruments.

In the context of bootstrap applications, further research will be focused on three aspects. The first concerns bootstrap distribution of LIML and Fuller estimators under weak instrument asymptotics. Applying residual–based resampling methods of Section 2.4, we evidence that they have a non–normal limiting distribution, presenting random quantities conditionally on the data D_T as viewed in the IV/TSLS case. In order to prove formally this fact, an open issue is related to the relationship between estimated $\hat{\kappa}$ and its bootstrapped counterpart, computed through bootstrap sample: $(y_t^*, x_t^*, Z_t^*)'$.

A second topic concerns new concepts of bootstrap validity, introduced in Cavaliere and Georgiev (2018), denoted as conditional on the sample and "on average" validity, which will be investigated in $\kappa-$class estimators and associated t/Wald tests. This could be done using different asymptotics to model $\boldsymbol{\pi}$, like the one proposed by Andrews and Chen (2012), called "semi–strong" instruments. Furthermore, we will be interested in the bootstrapped distribution of IV/TSLS estimators under weakly endogenous instruments. In this situation, the presence of a non–zero correlation between instruments and structural disturbances may be reflected in the bootstrap world, as suggested by our empirical findings regarding rejection frequeny of bootstrapped t-statistic in Section 3.3.4.

Finally, in further research, we will apply our bootstrap–based test in case of multiple endogenous regressors, using TSLS and other $\kappa-$class estimators. These models are rarely present in applied work, essentially because detection of weak instruments is still of practical interest when $m > 1$ and existing methods remain under debate. The choice of bootstrap replications $\overline{B}$ will play also a central role in further works, in order to improve the performance of proposed test in terms of empirical size and power.

Bibliography

D. Acemoglu, S. Johnson, and J. A. Robinson. The colonial origins of comparative development: An empirical investigation. *American economic review*, 91(5):1369–1401, 2001.

D. Acemoglu, S. Johnson, J. A. Robinson, and P. Yared. Income and democracy. *American Economic Review*, 98(3):808–42, 2008.

D. Andrews and J. H. Stock. Inference with weak instruments, 2005.

D. W. Andrews. Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica*, 68(2):399–405, 2000.

I. Andrews and A. Mikusheva. Weak identification in maximum likelihood: A question of information. *American Economic Review*, 104(5):195–99, 2014.

I. Andrews, J. Stock, and L. Sun. Weak instruments in iv regression: Theory and practice. *Working paper*, 2018.

G. Angelini, G. Cavaliere, and L. Fanelli. Bootstrapping dsge models. *Working Paper, Quaderni di Dipartimento Dipartimento di Scienze Statistiche "Paolo Fortunati" Serie Ricerche (3). ISSN 1973-9346*, 2016.

J. D. Angrist and A. B. Kreueger. Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4):979–1014, 1991.

J. D. Angrist and J.-S. Pischke. Mostly harmless econometrics: An empiricistvs companion. *Princeton Univ Pr*, 2008.

J. D. Angrist, G. W. Imbens, and A. B. Krueger. Jackknife instrumental variables estimation. *Journal of Applied Econometrics*, 14(1):57–67, 1999.

R. Ashley. Assessing the credibility of instrumental variables inference with imperfect instruments via sensitivity analysis. *Journal of Applied Econometrics*, 24(2):325–337, 2009.

I. V. Basawa, A. K. Mallik, W. P. McCormick, J. H. Reeves, and R. L. Taylor. Bootstrapping unstable first-order autoregressive processes. *The annals of Statistics*, 19 (2):1098–1101, 1991.

P. A. Bekker. Alternative approximations to the distributions of instrumental variable estimators. *Econometrica: Journal of the Econometric Society*, pages 657–681, 1994.

S. Blomquist and M. Dahlberg. Small sample properties of liml and jackknife iv estimators: Experiments with weak instruments. *Journal of Applied Econometrics*, pages 69–88, 1999.

C. Bontemps and N. Meddahi. Testing normality: a gmm approach. *Journal of Econometrics*, 124(1):149–186, 2005.

J. Bound, D. A. Jaeger, and R. M. Baker. 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(430): 443–450, 1995.

M. Bun and M. de Haan. Weak instruments and the first stage f-statistic in iv models with a nonscalar error covariance structure. *UvA-Econometrics Working Paper*, pages 10–02, 2010.

L. Camponovo and T. Otsu. Breakdown point theory for implied probability bootstrap. *The Econometrics Journal*, 15(1):32–55, 2012.

M. Caner. A pretest to differentiate between weak and nearly-weak instrument asymptotics. *International Econometric Review (IER)*, 3(2):13–21, 2011.

F. Canova and L. Sala. Back to square one: Identification issues in dsge models. *Journal of Monetary Economics*, 56(4):431–449, 2009.

A. J. Canty, A. C. Davison, D. V. Hinkley, and V. Ventura. Bootstrap diagnostics and remedies. *Canadian Journal of Statistics*, 34(1):5–27, 2006.

D. Card. Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, 69(5):1127–1160, 2001.

G. Cavaliere and I. Georgiev. Inference under random limit bootstrap measures. *Working Paper*, 2018.

G. Cavaliere, I. Georgiev, and A. R. Taylor. Unit root inference for non-stationary linear processes driven by infinite variance innovations. *Econometric Theory*, pages 1–47, 2016.

Y. Chang and J. Y. Park. A sieve bootstrap for the test of a unit root. *Journal of Time Series Analysis*, 24(4):379–400, 2003.

T. W. Chau. On the equivalence of indirect inference and bootstrap bias correction for linear iv estimators. *Economics Letters*, 123(3):333–335, 2014.

T. G. Conley, C. B. Hansen, and P. E. Rossi. Plausibly exogenous. *Review of Economics and Statistics*, 94(1):260–272, 2012.

J. G. Cragg and S. G. Donald. Testing identifiability and specification in instrumental variable models. *Econometric Theory*, 9(2):222–240, 1993.

R. Davidson and J. G. MacKinnon. Improving the reliability of bootstrap tests with the fast double bootstrap. *Computational Statistics & Data Analysis*, 51(7):3259–3281, 2007.

R. Davidson and J. G. MacKinnon. Bootstrap inference in a linear equation estimated by instrumental variables. *The Econometrics Journal*, 11(3):443–477, 2008.

R. Davidson and J. G. MacKinnon. Wild bootstrap tests for iv regression. *Journal of Business & Economic Statistics*, 28(1):128–144, 2010.

R. Davidson and J. G. MacKinnon. Bootstrap confidence sets with weak instruments. *Econometric Reviews*, 33(5-6):651–675, 2014.

R. Davidson and J. G. Mackinnon. Bootstrap tests for overidentification in linear regression models. *Econometrics*, 3(4):825–863, 2015.

F. Doko Tchatoka. On bootstrap validity for specification tests with weak instruments. *The Econometrics Journal*, 18(1):137–146, 2015.

S. G. Donald and W. K. Newey. Choosing the number of instruments. *Econometrica*, 69(5):1161–1191, 2001.

J. A. Doornik and H. Hansen. An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, 70:927–939, 2008.

P. Dovonon and S. Gonçalves. Bootstrapping the gmm overidentification test under first-order underidentification. *Journal of Econometrics*, 201(1):43–71, 2017.

J.-M. Dufour. Identification, weak instruments, and statistical inference in econometrics. *Canadian Journal of Economics/Revue canadienne d'économique*, 36(4): 767–808, 2003.

K. Finlay and L. M. Magnusson. Bootstrap methods for inference with cluster sample iv models. *Working Paper Available at SSRN 2574521*, 2014.

A. Flores-Lagunes. Finite sample evidence of iv estimators under weak instruments. *Journal of Applied Econometrics*, 22(3):677–694, 2007.

D. Freedman et al. On bootstrapping two-stage least-squares estimates in stationary linear models. *The Annals of Statistics*, 12(3):827–842, 1984.

F. A. R. Gomes and L. S. Paz. Narrow replication of yogo (2004) estimating the elasticity of intertemporal substitution when instruments are weak. *Journal of Applied Econometrics*, 26(7):1215–1216, 2011.

W. H. Greene. *Econometric analysis (International edition)*. Pearson US Imports & PHIPEs, 2000.

Z. Grichiles. Estimating the returns to schooling: Some econometrics problems. *Econometrica*, 45(1):1–22, 1977.

J. Hahn and J. Hausman. A new specification test for the validity of instrumental variables. *Econometrica*, 70(1):163–189, 2002a.

J. Hahn and J. Hausman. Notes on bias in estimators for simultaneous equation models. *Economics Letters*, 75(2):237–241, 2002b.

J. Hahn and J. Hausman. Weak instruments: Diagnosis and cures in empirical econometrics. *American Economic Review*, 93(2):118–125, 2003.

J. Hahn and A. Inoue. A monte carlo comparison of various asymptotic approximations to the distribution of instrumental variables estimators. *Econometric Reviews*, 21 (3):309–336, 2002.

J. Hahn, J. Hausman, and G. Kuersteiner. Estimation with weak instruments: Accuracy of higher-order bias and mse approximations. *The Econometrics Journal*, 7(1): 272–306, 2004.

A. R. Hall, G. D. Rudebusch, and D. W. Wilcox. Judging instrument relevance in instrumental variables estimation. *International Economic Review*, pages 283–298, 1996.

Z. Hanusz, J. Tarasinska, and W. Zielinski. Shapiro-wilk test with known mean. *REVSTAT-Statistical Journal*, 14(1):89–100, 2016.

J. Hausman, J. H. Stock, and M. Yogo. Asymptotic properties of the hahn–hausman test for weak-instruments. *Economics Letters*, 89(3):333–342, 2005.

J. A. Hausman, W. K. Newey, T. Woutersen, J. C. Chao, and N. R. Swanson. Instrumental variable estimation with heteroskedasticity and many instruments. *Quantitative Economics*, 3(2):211–255, 2012.

J. L. Horowitz. The bootstrap. In *Handbook of econometrics*, volume 5, pages 3159–3228. Elsevier, 2001.

Y.-S. Hsu, K.-N. Lau, H.-G. Fung, and E. F. Ulveling. Monte carlo studies on the effectiveness of the bootstrap bias reduction method on 2sls estimates. *Economics Letters*, 20(3):233–239, 1986.

A. Inoue. A bootstrap approach to moment selection. *The Econometrics Journal*, 9 (1):48–75, 2006.

T. Kitagawa. A test for instrument validity. *Econometrica*, 83(5):2043–2063, 2015.

F. Kleibergen. Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica*, 70(5):1781–1803, 2002.

F. Kleibergen and S. Mavroeidis. Weak instrument robust tests in gmm and the new keynesian phillips curve. *Journal of Business & Economic Statistics*, 27(3):293–311, 2009.

F. Kleibergen and R. Paap. Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, 133(1):97–126, 2006.

S. D. Levitt. Using electoral cycles in police hiring to estimate the effects of police on crime: Reply. *American Economic Review*, 92(4):1244–1250, 2002.

R. Y. Liu et al. Bootstrap procedures under some non-iid models. *The Annals of Statistics*, 16(4):1696–1708, 1988.

E. Mammen. Bootstrap and wild bootstrap for high dimensional linear models. *The annals of statistics*, pages 255–285, 1993.

R. S. Mariano. Simultaneous equation model estimators: Statistical properties and practical implications. *A companion to theoretical econometrics*, pages 122–143, 2001.

M. J. Moreira. A conditional likelihood ratio test for structural models. *Econometrica*, 71(4):1027–1048, 2003.

M. J. Moreira, J. R. Porter, and G. A. Suarez. Bootstrap validity for the score test when instruments may be weak. *Journal of Econometrics*, 149(1):52–64, 2009.

A. L. Nagar. The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. *Econometrica: Journal of the Econometric Society*, pages 575–595, 1959.

C. Nelson and R. Startz. The distribution of the instrumental variables estimator and its t-ratio when the instrument is a poor one. *The Journal of Business*, 63(1): S125–40, 1990a.

C. R. Nelson and R. Startz. Some further results on the exact small sample properties of the instrumental variable estimator. *Econometrica: Journal of the Econometric Society*, pages 967–976, 1990b.

C. R. Nelson and R. Startz. The zero-information-limit condition and spurious inference in weakly identified models. *Journal of Econometrics*, 138(1):47–62, 2007.

J. L. M. Olea and C. Pflueger. A robust test for weak instruments. *Journal of Business & Economic Statistics*, 31(3):358–369, 2013.

P. Oreopoulos. Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review*, 96(1): 152–175, 2006.

R. Ouysse. Computationally efficient approximation for the double bootstrap mean bias correction. *Economics Bulletin*, 31(3):2388–2403, 2011.

R. Ouysse. A fast iterated bootstrap procedure for approximating the small-sample bias. *Communications in Statistics-Simulation and Computation*, 42(7):1472–1494, 2013.

J.-S. Pischke and T. Von Wachter. Zero returns to compulsory schooling in germany: Evidence and interpretation. *The Review of Economics and Statistics*, 90(3):592–598, 2008.

D. S. Poskitt, C. Skeels, et al. Inference in the presence of weak instruments: A selected survey. *Foundations and Trends® in Econometrics*, 6(1):1–99, 2013.

R Core Team. R: A language and environment for statistical computing, 2016. URL `https://www.R-project.org/`.

T. J. Rothenberg. Approximating the distributions of econometric estimators and test statistics. *Handbook of econometrics*, 2:881–935, 1984.

E. Sanderson and F. Windmeijer. A weak instrument f-test in linear iv models with multiple endogenous variables. *Journal of Econometrics*, 190(2):212–221, 2016.

J. Shea. Instrument relevance in multivariate linear models: A simple measure. *Review of Economics and Statistics*, 79(2):348–352, 1997.

C. Skeels and F. Windmeijer. On the stock-yogo tables. Technical report, Department of Economics, University of Bristol, UK, 2016.

D. Staiger and J. H. Stock. Instrumental variables regression with weak instruments. *Econometrica*, 65(3):557–586, 1997.

J. Stock and M. Yogo. *Asymptotic distributions of instrumental variables statistics with many instruments*, volume 6. Chapter, 2005.

J. H. Stock, J. H. Wright, and M. Yogo. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, 20(4):518–529, 2002.

W. Wang and M. Kaffo. Bootstrap inference for instrumental variable models with many weak instruments. *Journal of Econometrics*, 192(1):231–268, 2016.

W. Wang, Q. Liu, et al. Bootstrap-based selection for instrumental variables model. *Economics Bulletin*, 35(3):1886–1896, 2015.

J. M. Wooldridge. *Econometric analysis of cross section and panel data*. MIT press, 2010.

C.-F. J. Wu. Jackknife, bootstrap and other resampling methods in regression analysis. *the Annals of Statistics*, pages 1261–1295, 1986.

M. Yogo. Estimating the elasticity of intertemporal substitution when instruments are weak. *Review of Economics and Statistics*, 86(3):797–810, 2004.

A. Young. Consistency without inference: Instrumental variables in practical application. *Unpublished manuscript, London: London School of Economics and Political Science. Retrieved from: http://personal. lse. ac. uk/YoungA*, 2017.

Z. Zhan. Detecting weak identification by bootstrap. *Working Paper*, 2017.

E. Zivot, R. Startz, and C. R. Nelson. Valid confidence intervals and inference in the presence of weak instruments. *International Economic Review*, pages 1119–1144, 1998.

List of Figures

List of Tables