University of Kent

School of Economics Discussion Papers

Regularization Based Anderson Rubin Tests for Many Instruments

Marine Carrasco and Guy Tchuente

September 2016

KDPE 1608

Regularization Based Anderson Rubin Tests for Many Instruments*

Marine Carrasco
University of Montreal

Guy Tchuente†
University of Kent

September 2016

Abstract

This paper studies the asymptotic validity of the regularized Anderson Rubin (AR) tests in linear models with large number of instruments. The regularized AR tests use information-reduction methods to provide robust inference in instrumental variable (IV) estimation for data rich environments. We derive the asymptotic properties of the tests. Their asymptotic distribution depend on unknown nuisance parameters. A bootstrap method is used to obtain more reliable inference. The regularized tests are robust to many moment conditions in the sense that they are valid for both few and many instruments, and even for more instruments than the sample size. Our simulations show that the proposed AR tests work well and have better performance than competing AR tests when the number of instruments is very large. The usefulness of the regularized tests is shown by proposing confidence intervals for the Elasticity of Intertemporal Substitution (EIS).

Keywords: Many weak instruments; AR test; Bootstrap; Factor Model.

*The authors thank the participants to the European winter meeting of the econometric society 2015, CEA 2016 for helpful comments. Carrasco gratefully acknowledges financial support from SSHRC.

†School of Economics and MaGHiC, Email: g.tchuente@kent.ac.uk

Non-technical summary

In many empirical works in economics, the aim of the researcher is to establish a causal or a noncausal relationship between two variables. Because unobserved variables affect most economics variables, identification and estimation of parameters of interest suffer from the endogeneity problem. In presence of endogeneity, identification of the causal parameter of interest is achieved using instrumental variables. The instrumental variables are assumed to be highly correlated with the right-hand side endogenous variables (strong) and uncorrelated with the structural error (valid or respecting the exclusion restriction).

In empirical applications, finding a valid instrumental variable that is at the same time strong is very difficult. This difficulty to have perfect instruments has encouraged the study of inference and estimation in presence of weak (weakly correlated with the right-hand side endogenous variables) instruments. We assume, in the present paper, that we have many valid (but potentially weak) instrumental variables. We use them along side with robust to weak instruments robust statistics to improve the quality of the inference. The Anderson-Rubin (AR) (Anderson and Rubin (1949)) test is an example of such robust to weak identification procedure.

We examine the regularization of the AR test when the number of instruments is large. The regularized AR tests use information-reduction methods to provide robust inference in instrumental variable estimation for data-rich environments. We derive the asymptotic properties of the regularized AR tests. Their asymptotic distributions depend on unknown eigenvalues and a regularization parameter. A bootstrap method is used to obtain more reliable inference. The regularized tests are robust to many moment conditions in the sense that they are valid for both few and many instruments, and even for more instruments than the sample size. We perform a limited set of Monte Carlo experiments. Our simulations show that the proposed AR tests work well and have better performance than competing AR tests when the number of instruments is very large.

1 Introduction

Identification and estimation of coefficients on endogenous variables in linear structural equations is the focus of many applied economics papers. The identification of these coefficients is achieved using instrumental variables that are assumed to be highly correlated with the right hand side endogenous variables (strong) and uncorrelated with the structural error (valid or respecting the exclusion restriction). When these conditions are respected, there are many estimation and inference procedures, such as two stages least squares (2SLS), limited information maximum likelihood (LIML) and full information maximum likelihood, that can be used to estimate the parameters and provide asymptotically valid inference. However, when the instruments are weak, ie the correlation between the endogenous variables and instruments is low, conventional asymptotics may provide poor approximations to the finite sample distributions of conventional estimators and test statistics. Using many valid (but potentially weak) instrumental variables, along side with robust to weak instruments robust statistics, can improve the quality of the inference.[1] The Anderson-Rubin (AR) (Anderson and Rubin (1949)) test is an example of such robust to weak identification procedure. This paper examines the regularization of AR test when the number of instrument is large.

An important problem with using many instrumental variables is that the conventional asymptotic behavior of the IV estimator and various test statistics deteriorate. A sizable literature on robust methods for many IVs inference currently exists. This literature considers estimation and inference in the presence of many (possibly weak) instruments (see, among others,Chao and Swanson (2005), Hansen, Hausman, and Newey (2008), Andrews and Stock (2007)). Andrews and Stock (2007) and Newey and Windmeijer (2009) show that the AR test statistic remains valid when the number of instruments grows at a slower rate than the sample size. However, finite sample studies suggest that test performance is highly sensitive to the number of instruments. More recently, Anatolyev and Gospodinov (2011) have studied the AR statistics under the many instruments assumption of Bekker (1994). This means that they allow the sample size and the number of instruments to grow at the same rate. They find that the asymptotic size of the standard AR test exceeds the nominal level when there are many instruments. They proposed a modification of the conventional AR tests that is based on critical values of a chi-squared distribution. The proposed corrected test is robust to the

[1]A large number of instruments can be constructed by interacting different variables (see Angrist and Krueger (1991)) or using lagged dependent variables in panel data models (see Arellano and Bond (1991)).

number of the moment conditions increase. However, it also suffers from size distortion when the number of instruments is very large and cannot be used when there is an infinite number of instruments or a continuum as in Carrasco (2012) and Carrasco and Tchuente (2015). Kapetanios, Khalaf, and Marcellino (2015) have also proposed a factor-based modifications of three popular weak-instrument robust statistics, AR, KLM (Kleibergen (2002)) and LR (Moreira (2003)). For the AR test statistic, they provided analytical finite sample results under the usual assumptions, in a factor model framework. However, they did not address the issue of the choose of the number of factors. Dufour and Valéry (2016) have used regularization methods for Wald-type tests in the presence of a possibly singular covariance matrix. The regularization methods used in their paper are spectral cut-off, Tikhonov and Landweber-Fridman regularization. They show that the asymptotic distribution of the Wald test can be simulated or bounded by simulation. Our work can be view as an extension of the use of regularization to AR test with the aim of correcting the many instruments problem.

This paper extends Anatolyev and Gospodinov (2011) and Kapetanios, Khalaf, and Marcellino (2015) by proposing a regularized version of the AR statistic. The regularisation is done using three methods. The first one is the Tikhonov regularization, the second is based on an iterative method called Landweber-Fridman and the third is based on the principal components associated with the largest eigenvalues (this regularized AR test corresponds to Kapetanios, Khalaf, and Marcellino (2015) factor based AR test). Asymptotic behaviors of the tests statistics are derived. The regularized AR tests statistic depends on the regularization parameter as well as unknown eigenvalues. To improve the regularized tests small sample properties, we propose a restricted efficient bootstrap test for inference. A Monte Carlo experiment reveals that the regularized AR tests perform well and are better than the Anatolyev and Gospodinov (2011) corrected AR test when the number of instruments is very large. More precisely, the regularized AR circumvent the size problems resulting from many instruments and improve the power of the AR test statistic. An empirical application proposes confidence intervals for the Elasticity of Intertemporal Substitution (EIS). It suggests that the EIS is less than one.

The remaining of the paper is organized as follows. Section 2 introduces the AR regularized test for continuum of moments. Section 3 proposes a bootstrap strategy. Section 4 presents Monte Carlo experiments Section 5 discusses inference on the value of the EIS and Section 6 concludes.

2 Model, assumptions and tests

This section presents the model used for our weak identification robust inference. It also discusses issues related to inference in the presence of many instruments using the AR test. We propose a regularized AR test. Under conventional assumptions, we derive the asymptotic distribution of the proposed regularized AR test.

2.1 Model

The structural equation is the following: The model is

$$y_i = W_i'\delta_0 + \varepsilon_i \tag{1}$$

$i = 1, 2, ..., n.$, $E(W_i\varepsilon_i) \neq 0$ which means that W_i is endogenous. y_i is a scalar and x_i is a vector of exogenous variables such that $E(\varepsilon_i|x_i) = 0$, we also assume conditional homoscedasticity $E(\varepsilon_i^2|x_i) = \sigma_\varepsilon^2 > 0$.

Estimation and inference on δ_0 can be carry out using the moment condition

$$E\left[(y_i - W_i'\delta_0)z_i\right] = 0$$

where z_i is a $1 \times L$ vector function of x_i. It can take the following form

- $z_i = x_i$ where x_i is a L- vector with a fixed L
- $z_{ij} = (x_i)^{j-1}$ with $j \in \mathbb{N}$, thus we have an infinite countable instruments.
- $z_i = exp(i\tau' x_i)$ where $\tau \in \mathbb{R}^{dim(x_i)}$, thus we have a continuum of moments.

The main focus is inference on the $p \times 1$ vector δ_0 when instruments are possibly weak (meaning that the correlation between W_i and z_i is low).
The proposed model is set up in a general framework that enables us to deal with a finite number of moments, a countable infinite number of moments, continuum of moments as well as factor model representation.

2.2 Conventional AR test

Since Dufour (1997), it is known that if the parameter set is allowed to include values where the model is not identified, then the correct confidence interval for a structural parameter must be unbounded with positive probability. Hence, bounded confidence intervals, such as Wald intervals formed in the usual way, are not correct. To overcome this inference problem, many robust to weak instruments methods have been proposed in the literature. The main tests are the likelihood ratio (LR), the Anderson-Rubin (AR), and Lagrange multiplier (LM) test statistics.

Let us focus our attention on the AR test. The number of instruments is L. A we wish to test the null hypothesis $H_0 : \delta = \delta_0$.

The AR test statistic is given by

$$AR(\delta_0) = \frac{n-L}{L} \frac{(y - W\delta_0)' P_Z (y - W\delta_0)}{(y - W\delta_0)'[I_n - P_Z](y - W\delta_0)},$$

where $P_Z = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'$ and $\mathbf{Z} = (z_1, z_2, ..., z_n)'$ is a $n \times L$ matrix of instruments.
The AR statistic possesses some appealing robustness properties (see Dufour and Taamouti (2007) for details on these properties) and is asymptotically distributed as $\chi^2(L)/L$.

The asymptotics of this test, as well as other robust to weak identification tests, have been studied when the number of instruments is large and the coefficients on the instruments are relatively small (Andrews and Stock (2007)). In these models, the number of instruments increases moderately ie the number of instruments grows asymptotically but slowly relative to the sample size. In their framework of moderately many instruments (more precisely, when $\frac{L^3}{n} \to 0$ as $L, n \to \infty$), Andrews and Stock (2007) show that:

$$\sqrt{L}(AR(\delta_0) - 1) \to \mathcal{N}(0, 2) \tag{2}$$

As pointed out by Anatolyev and Gospodinov (2011), the asymptotic result in (2) is not valid under Bekker (1994) asymptotic as $\frac{L}{n} \to c$ where c is a constant. Anatolyev and Gospodinov provided the corrected critical values for the conventional AR test statistic. They showed that

$$\sqrt{L}(AR(\delta_0) - 1) \to \mathcal{N}(0, 2/(1-\lambda))$$

when $\frac{L}{n} \to \lambda$ as $L, n \to \infty$,

This asymptotic distribution covers many situations in applied work. However, the asymptotic distribution is very flat for λ close to 1. Moreover, this test cannot be used when $L > n$ or when there is a continuum of instruments, see Carrasco and Florens (2000), Carrasco, Florens, and Renault (2007) because $\mathbf{Z}'\mathbf{Z}$ is singular. The next section develops a regularized AR test that can be used to overcome these limitations.

2.3 Regularized AR test with many instruments

This section proposes a regularized version of the AR statistic to test: $H_0 : \delta = \delta_0$.

This regularized AR statistic is a modified version of the classic AR statistic which is robust to weak instruments. The conventional AR test provided in the literature suffers from size distortion when many instruments are used. A weakness shared by many other robust to weak identification test. Alternative asymptotics have been considered but they can only be used in the case with a countable number of instruments and with a number of instruments less than sample size ($L < n$).

The regularized Anderson Rubin statistic can be used in the case of many instruments and also in the case of a continuum of instruments.

Let $\mathbf{Z}$ denote the $n \times L$ matrix having columns given by z_i. ψ_j denote the orthonormal eigenvectors of the $n \times n$ matrix $\frac{\mathbf{ZZ}'}{n}$ associated with eigenvalues λ_j such that

$$\frac{\mathbf{ZZ}'}{n}\psi_j = \sqrt{\lambda_j}\psi_j$$

Recall that AR test statistic involves a projection matrix

$$P_Z = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'.$$

The matrix $\mathbf{Z}'\mathbf{Z}$ may become nearly singular when L gets large. Moreover, $\mathbf{Z}'\mathbf{Z}$ is singular whenever $L \geq n$. In order to investigate all these possibilities, we will consider a regularized version of the inverse of the matrix $\mathbf{Z}'\mathbf{Z}$.

For an arbitrary $n \times 1$ vector v, we define the $n \times n$ matrix P^α as

$$P^\alpha v = \frac{1}{n}\sum_{j=1}^{n} q\left(\alpha, \lambda_j^2\right)\left(v'\psi_j\right)\psi_j \tag{3}$$

where $q\left(\alpha, \lambda_j^2\right)$ is a weight that takes different forms depending on the regularization schemes and α is the regularization parameter. We consider four types of regularization:

- The Tikhonov (T) regularization: $q\left(\alpha, \lambda_j^2\right) = \frac{\lambda_j^2}{\lambda_j^2 + \alpha}$.
- The Landweber-Fridman (LF) regularization: $q(\alpha, \lambda_j^2) = [1 - (1 - c\lambda_j^2)^{1/\alpha}]$, where c is a constant such that $0 < c < 1/\left\|\mathbf{Z}'\mathbf{Z}/n\right\|^2$ and $\left\|\mathbf{Z}'\mathbf{Z}/n\right\|$ denotes the largest eigenvalue of $\mathbf{ZZ}'/n$.
- Spectral cut-off (SC): $q(\alpha, \lambda_j^2) = I(\lambda_j^2 \leq \alpha)$. Only eigenvector larger than a threshold fixed by the researcher are consider.

- The Principal Components (PC): $q(\alpha, \lambda_j^2) = I(j \leq 1/\alpha)$ where $1/\alpha$ is a positive integer.

PC and SC are equivalent if the eigenvalue are all of multiplicity 1.

Note that all these regularization techniques involve a tuning parameter α. The case $\alpha = 0$ corresponds to the case without regularization, $q\left(\alpha, \lambda_j^2\right) = 1$. Then, we obtain

$$P^0 = P_Z = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'.$$

The regularized AR test statistic is the following:

$$AR_R(\delta_0) = \frac{n(y - W\delta_0)'P^\alpha(y - W\delta_0)}{(y - W\delta_0)'[I_n - P^\alpha](y - W\delta_0)}$$

where P^α is the regularized version[2] of the projection matrix on the space spanned by the instrumental variables.

Assumption 1. The elements of $\{y_i, W_i, x_i\}_{i=1,\ldots n}$ are *iid*.

Assumption 1 means that we have an independent identically distributed sample of the population. By assuming *iid* data we are restricting our result to the cases where there is correlation or spatial correlation among observations. The assumption is proposed mainly to simplify the problem in order to focus our attention on the many instruments problem on AR test.

In the case where we have an infinite number of instruments, there is a need to define covariance operator. The covariance operator is the infinite dimensional counterpart of $\mathbf{Z}'\mathbf{Z}$. We denote the covariance operator of the instruments by $\mathbf{K}$.

Assumption 2. The operator $\mathbf{K}$ is trace-class or nuclear.

The trace-class property of the covariance operator insures us of the possibility to use a countably infinite number of eigenvalues and orthonormal eigenvectors to represent the operator.

The following theorem establishes the asymptotic behavior of the regularized AR_R test under the null.

[2]Appendix A gives a detail definition of the regularization methods and the generalization of P^α to the case with continuum of moment conditions.

Theorem 1. *For a fixed value of α, under Assumptions 1, 2 and the null hypothesis $H_0 : \delta = \delta_0$;*

$$AR_R(\delta_0) - J_0 = o_p(1)$$

as n goes to infinity with $J_0 = \dfrac{\varepsilon' P^\alpha \varepsilon}{\sigma_\varepsilon^2} \xrightarrow{d} \chi \equiv \sum_{j=1}^{\infty} q_j \chi_j^2(1)$.

Where $\chi_j^2(1)$ denote independent chi-square with 1 degree of freedom random variables.

Proof: In appendix B

The asymptotic distribution of our test statistic depends on unknown eigenvalues and on a regularized parameter that is fixed. The asymptotic distribution is an unknown sum of known object. It can be approximated by a finite weighted sum of chi-square. However, the approximation of the asymptotic distribution may differ from its true value. The difference can be caused by the number of terms in the sum.[3] The two distributions can also differ because of the estimates of the eigenvalues. In practice, the eigenvalues use in the weight need to be estimated, if the estimators are consistent then the estimation of the eigenvalues is not a problem. The performance of the test also depends on the regularization parameter.

Theorem 1 discusses the general regularized AR tests in the cases where the number of instruments can be infinite, finite or a continuum. The following corollary derives the asymptotic behavior of the AR tests when the number of instruments is large but finite.

Corollary 1. *(with a finite number of instruments, $L < n$) For fixed values of α and L,under Assumption 1 and the null hypothesis $H_0 : \delta = \delta_0$;*

$$J_0 = \frac{\varepsilon' P^\alpha \varepsilon}{\sigma_\varepsilon^2} \xrightarrow{d} \chi \equiv \sum_{j=1}^{L} q_j \chi_j^2(1).$$

with $\chi_j^2(1)$ denoting independent chi-square with 1 degree of freedom random variables.

The number of instruments is finite (L), which gives an asymptotic distribution that is a sum of L weighted independent $\chi^2(1)$. The weight depends on the regularization technique used. The regularization methods all involve a regularization parameter, that we have assumed in this paper to be fixed. As the asymptotic distribution of the AR tests statistic depends on the regularization parameter, their ability to control for size will be affected by it.

[3]The investigation of the property of the approximation and the number of terms to be used is left for future research.

In the case with a finite number of instruments, as well as, in the case with an infinite or continuum of instruments, the asymptotic distribution of the test depends on eigenvalues (to be estimated in practice) and the regularized parameter. To mitigate the effect of both the eigenvalues estimation and the choice of the regularization parameter on the inference, we used a bootstrap test strategy. The use of a bootstrap allows us to avoid the estimation of the eigenvalues. Moreover, good asymptotic results can be obtained using any regularization parameter.

The following remarks discuss a case of regularized AR test in a model with a factor structure. We apply corollary 1 to these model using PC regularization this shows the links between our work and Kapetanios, Khalaf, and Marcellino (2015).

Remark 1: For the PC regularization method, if the number of principal components chosen is r, the asymptotic distribution is simple enough to allow direct inference.

$$AR_R(\delta_0) \xrightarrow{d} \chi \equiv \sum_{j=1}^{\infty} q_j \chi_j^2(1) = \chi^2(r),$$

We consider the following DGP coming from a factor model.

$$y_i = W_i'\delta_0 + \varepsilon_i \quad (4)$$

$$W_i = F_i\pi + v_i \quad (5)$$

$$x_i = F_i\Delta + e_i \quad (6)$$

The specificity of this DGP is that d the endogenous variables W_i depend on r unobservable, independent factors $f_i = (f_{i1}, ..., f_{ir})$. Each element of of the exogenous variable x_i depends on the common factors f_i via the loadings Δ, and on an idiosyncratic component e_i.

Kapetanios, Khalaf, and Marcellino (2015) show that, under standard assumption on the errors term and factor structure, a slightly modified version of the AR statistics converges to $\chi^2(r)$. The modified AR in their case is very similar to the regularized AR using PC regularisation, in particular in the cases where factor are estimated using principal component method.

In their framework, the number of factors is known. However, in reality factors are unobservable and need to be estimated. Remark 2 discusses the cases in which the estimated number of factor is biased and how the inference using Kapetanios, Khalaf, and Marcellino (2015) asymptotic is affected.

We can also note that result of Theorem 1 is an extension of Kapetanios, Khalaf, and Marcellino (2015)'s result to different regularization methods in order to deal with the large number of instruments. Moreover or result is not restricted to factor model, it also applies to infinite or continuum of instruments.

Remark 2: Let us assume that there are r factors. In practice this number of factor is unobserved and unknown, the econometricians would used r_1 factors.

If we use r_1 PC with $r_1 < r$, then $AR_R(\delta_0) \xrightarrow{d} \chi \equiv \sum_{j=1}^{\infty} q_j \chi_j^2(1) = \chi^2(r_1)$, the PC regularized AR test rejects the null hypothesis less frequently and the test is liberal.
If we use r_1 PC with $r_1 > r$, then $AR_R(\delta_0) \xrightarrow{d} \chi \equiv \sum_{j=1}^{\infty} q_j \chi_j^2(1) = \chi^2(r_1)$, the PC regularized AR test rejects the null hypothesis more frequently and the test is conservative.

The above remark implies that the choice of number of the factors can affect the reliability of the inferences using PC regularized AR. These results are in line with Dufour and Valéry (2016) who have found in thier simulations a loss of power (in certain directions) for the spectral cut-off Wald statistic.

In situations where there is no factor structure in the data, other regularization methods can be used as dimensional reduction tools. The use of the method could lead better results as in the case forecasting of misspecified factor model investigated by Carrasco and Rossi (2016).

Note that we can construct confidence intervals for the joint parameter vector δ_0 by inverting the $AR_R(\delta_0)$. Specifically, let $q_{\kappa,\alpha}$ be the $1-\kappa$ quantile of the asymptotic distribution of χ. The $1-\kappa$ confidence interval is

$$\{\delta, AR_R(\delta) \leq q_{\kappa,\alpha}\}.$$

This confidence interval will be asymptotically valid under weak identification and for many weak instruments. Individual confidence intervals are obtained from the joint confidence interval by projection.

In the case with finite many instruments, the asymptotic distribution of our test statistic is simple enough to enable simulated asymptotic distribution. The asymptotic distribution can be simulated for a fixed value of α, and critical value is obtained by computing the corresponding quantile of the simulated distribution.

However, in the general case of an infinite or a continuum of instruments these critical values are very difficult to compute. Furthermore, the regularized statistic proposed depends on unknown eigenvalues. In order to made reliable inference we develop a bootstrap procedure.

3 Bootstrapped regularized AR tests

In recent years, the use of bootstrapping to perform hypothesis testing in econometrics, has become common. Its use for testing purpose has been advocated by Hall and Horowitz (1996), Davidson and MacKinnon (1999), and several others. However, its use for inference in linear model estimation with instrumental variables has not been very popular and this may be because the simplest bootstrap methods for this problem do not work very well, see, for example, Flores-Lagunes (2007). Recently Davidson and MacKinnon (2008), Davidson and MacKinnon (2010) , Davidson and MacKinnon (2014a), Davidson and MacKinnon (2014b) and Wang and Kaffo (2014) have proposed more sophisticated bootstrap methods that work much better than traditional bootstrap procedures, even when they are combined with the usual t statistic. Davidson and MacKinnon (2008, 2010, 2014) study different type of bootstrap methods. Their simulation results show that the restricted efficient (RE) bootstrap[4] and wild restricted efficient bootstrap (WRE) approaches perform very well relative to other methods. Davidson and MacKinnon (2014a) propose a RE bootstrap confidence sets based on t statistics. They also show that the procedures that generally work best are CLR confidence sets using asymptotic critical values and bootstrap confidence sets based on LIML estimates. In a companion paper Davidson and MacKinnon (2014b), the AR confidence sets are investigated. Their results, however, show that AR confidence sets have many undesirable properties. But, the AR confidence sets have correct coverage under classical assumptions. In addition, the AR test possesses many robustness properties (eg: robust to instrument omission and to the misspecification of the reduced form, see Dufour and Taamouti (2007)). However, these papers focus on the case where the number of instruments is kept small relative to the sample size.

Wang and Kaffo (2014) study bootstrap-based inference methods with many weak instruments. They propose a modification to the RE bootstrap method and prove that it provides a valid distributional approximation for LIML with many weak instruments. They show that the modified bootstrap procedure has better performance relative to RE bootstrap method.

We propose a new RE bootstrap procedure. The procedure imposes the null hypotheses on the structural equation. An efficient estimation of the reduced form is obtained using regularization methods. These bootstrap regularized AR tests allow for many instruments and

[4] The bootstrap data are generated under the null (Restricted) and use efficient estimates of the reduced-form equation (Efficient).

continuum of instruments. This procedure solves the problem of the choice of the regularized parameter and alleviates the small sample size distortion of the regularized AR tests.

The bootstrap procedure we propose, to test the null hypothesis $H_0 : \delta = \delta_0$, follows four steps.

Step 1: Compute the re-centered residuals $\tilde{\varepsilon}_i = \hat{\varepsilon}_i - \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i$ and $\tilde{u}_i = \hat{u}_i - \frac{1}{n}\sum_{i=1}^{n}\hat{u}_i$ which are obtained from

$\hat{\varepsilon}_i = y_i - W_i\hat{\delta}$ and $\hat{u}_i = W_i - P^{\widetilde{\alpha}}W_i$,

$\hat{\delta}$ is estimated using regularized LIML of Carrasco and Tchuente (2015) and $\widetilde{\alpha}$ is the chosen regularization parameter in the estimation of δ, $\widetilde{\alpha}$ converges to zero.

Step 2: The bootstrap data-generating process (DGP) is obtained as follows

$\varepsilon_i^* = \tilde{\varepsilon}_i$ and $u_i^* = \tilde{u}_i$ with $\tilde{\varepsilon}_i$ and $\tilde{u}_i$ drawn with replacement from the re-centered residuals obtained in Step 1.

$$\begin{cases} y_i^* = W_i^{*\prime}\delta_0 + \varepsilon_i^* \\ W_i^* = P^{\widetilde{\alpha}}W_i + u_i^* \end{cases}$$

Step 3: By repeating Step 2, we generate B bootstrap samples indexed by b. From each sample, a bootstrap test statistic is computed

$$AR_{Rb}^* = \frac{n(y^* - W^*\delta_0)'P^{\alpha}(y^* - W^*\delta_0)}{(y^* - W^*\delta_0)'[I_n - P^{\alpha}](y^* - W^*\delta_0)}.$$

Step 4: We construct the bootstrap P-value:

$$\hat{p}^*(AR_R) = \frac{1}{B}\sum_{b=1}^{B} I(AR_{Rb}^* > AR_R)$$

Step 5: Reject the null hypothesis if $\hat{p}^*(AR_R) < \kappa$ where κ is the level of the test.

We make the following assumption.

Assumption 3: (i) $\tilde{\alpha}$ converges to zero.
(ii) $E(\|W_i\|^2) \leq \infty$ and $E(W_i|x_i)$ is in the space spanned by instrumental variables.

The following proposition shows the bootstrap validity of the AR_R statistic under many weak instruments assumptions. Let us define t

$$AR_{Rb}^* = \frac{n(y^* - W^*\delta_0)'P^{\alpha}(y^* - W^*\delta_0)}{(y^* - W^*\delta_0)'[I_n - P^{\alpha}](y^* - W^*\delta_0)}.$$

Theorem 2. *For a fixed value of α, under Assumptions 1-3. Then as n goes to infinity,*

$$AR_R^*(\delta_0) \xrightarrow{d^*} \chi \equiv \sum_{l=1}^{\infty} q_l\chi_k^2(1)$$

In addition, since the limiting distribution function is continuous,

$$\sup_{x\in\mathbf{R}} \left(|P^*(AR_R^*(\delta_0) \leq x) - P(AR_R(\delta_0) \leq x)|\right) \xrightarrow{p} 0$$

where P^* denotes the probability measure induced by the i.i.d. bootstrap and $\xrightarrow{d^*}$ denotes weak convergence under the bootstrap probability measure.

Proof In appendix.

This theorem shows that the bootstrap procedure proposed here is valid.

4 Simulation for Regularized AR test

To evaluate the finite-sample performance of the proposed tests, we conduct a small simulation study as in Anatolyev and Gospodinov (2011).

The data for the Monte Carlo experiment are generated from the model

$$y_i = W_i'\delta_0 + \varepsilon_i,$$

$$W_i = x_i'\pi + u_i,$$

Where $(\varepsilon_i, u_i)' = chol(\Sigma)\xi_i$ with $\xi_i \sim \mathcal{N}(0, I_2)$, $v_i \sim \mathcal{N}(0,1)$, $x_i \sim \mathcal{N}(0, I_L)$

$$\pi = \sqrt{\frac{1}{L}}\iota_L \quad and \quad \Sigma = \begin{pmatrix} 0.25 & 0.20 \\ 0.20 & 0.25 \end{pmatrix}$$

where ι_L is an L-vector of ones and $\lambda = L/n$.

Tables 1, 2 and 3 present the empirical size at 5% nominal level of AR, AR_{corr}, $AR_{R.}$ and $AR_{R.M}$ test which denote respectively the conventional AR test, the modified AR test proposed in Anatolyev and Gospodinov (2011), the regularized AR test proposed in this paper using bootstrap, and the regularized AR test using simulated critical values (T, P and L subscripts are respectively for Tikhonov, Principal component and Landweber-Fridman). These results are based on 1000 Monte Carlo replications. The purpose of this experiment is to compare the quality of the three approximations corresponding to different asymptotic frameworks. We consider values of $\lambda = \frac{L}{N}$ equal to 0.04, 0.2, 0.5, 0.8 and 1.1. The values of λ are used in combination with sample sizes of 100, 200 and 500. For the regularized AR using simulated Monte Carlo tests, the regularized values are fixed. For Tikhonov $\alpha = 25$, three first Principal Components are used and the number of iteration is five for Landweber-Fridman.

In the case of the bootstrap regularized AR, the same regularization parameter selected during the estimation step is used to construct the regularized AR. The optimal α for Tikhonov is searched over the interval [0.01, 0.5]. The range of values for the number of iterations for LF is from 1 to 100 and, for the number of principal components, it is from 1 to the number of instruments.

Table 1: Empirical level of Regularized Tikhonov AR tests; nominal level 0.05

N	λ	0.04	0.2	0.5	0.8	1.1
100	AR	0.062	0.076	0.135	0.274	NA
	AR_{corr}	0.06	0.052	0.068	0.095	NA
	AR_{RTM}	0.058	0.049	0.046	0.061	0.073
	AR_{RT}	0.057	0.042	0.057	0.06	0.067
200	AR	0.062	0.088	0.145	0.279	NA
	AR_{corr}	0.059	0.066	0.058	0.079	NA
	AR_{RTM}	0.046	0.048	0.043	0.068	0.085
	AR_{RT}	0.061	0.059	0.052	0.057	0.047
500	AR	0.047	0.071	0.115	0.25	NA
	AR_{corr}	0.042	0.04	0.05	0.077	NA
	AR_{RTM}	0.05	0.043	0.06	0.085	0.157
	AR_{RT}	0.046	0.044	0.039	0.039	0.05

NB: AR, AR_{corr}, AR_{RT} and AR_{RTM} denote the conventional AR test, the modified AR test proposed in Anatolyev and Gospodinov (2011), the regularized AR Tikhonov, test proposed in this paper using bootstrap, and the regularized AR Tikhonov test using simulated critical values, respectively.

Size results in Tables 1 to 3 suggest that there are considerable size distortions for the AR statistic when the number of instruments is large. The corrected Anatolyev and Gospodinov (2011)'s AR test reduces these distortions in large samples. However, the corrected AR test small sample performances are not very good. Both the AR and the AR_{corr} can not be computed when the number of instruments is larger than the sample size.

Table 1 suggests that the regularized Tikhonov AR has better size control for all numbers of instruments. The regularized Tihkonov AR test using the bootstrap deliver better best result, it is size correct in large samples ($N = 500$). The regularized Tikhonov tests with simulated critical values have better performance in small sample. However, the performance depends

Table 2: Empirical level of Regularized PC AR tests; nominal level 0.05

N	λ	0.04	0.2	0.5	0.8	1.1
100	AR	0.062	0.076	0.135	0.274	NA
	AR_{corr}	0.06	0.052	0.068	0.095	NA
	AR_{RPM}	0.068	0.054	0.056	0.046	0.056
	AR_{RP}	0.059	0.045	0.045	0.059	0.049
200	AR	0.062	0.088	0.145	0.279	NA
	AR_{corr}	0.059	0.066	0.058	0.079	NA
	AR_{RPM}	0.045	0.049	0.049	0.06	0.056
	AR_{RP}	0.061	0.057	0.044	0.044	0.059
500	AR	0.047	0.071	0.115	0.25	NA
	AR_{corr}	0.042	0.04	0.05	0.077	NA
	AR_{RPM}	0.047	0.055	0.052	0.052	0.055
	AR_{RP}	0.047	0.036	0.046	0.062	0.053

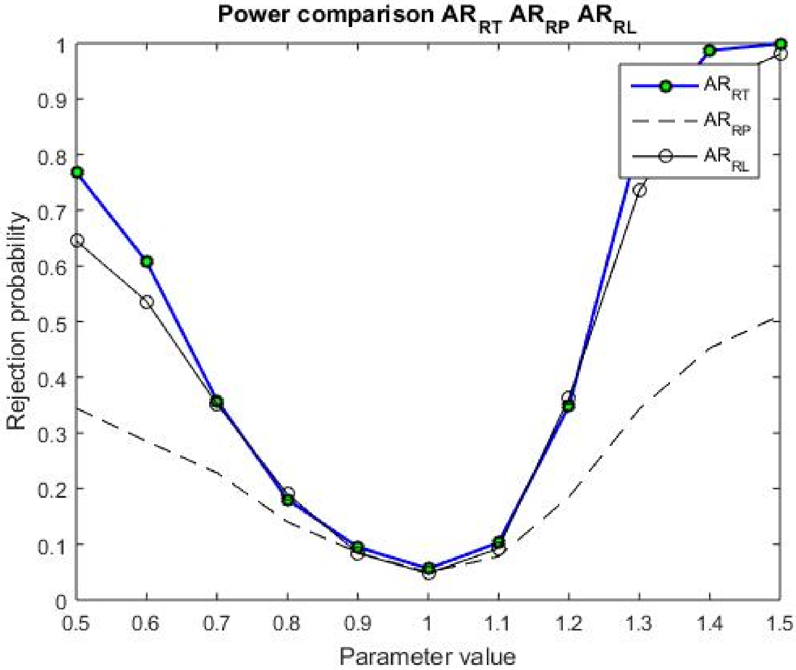
NB: AR, AR_{corr}, AR_{RP} and AR_{RPM} denote the conventional AR test, the modified AR test proposed in Anatolyev and Gospodinov (2011), the regularized AR Principal component test proposed in this paper using bootstrap, and the regularized AR Principal component test using simulated critical values, respectively.

Table 3: Empirical level of regularized LF AR tests; nominal level 0.05

N	λ	0.04	0.2	0.5	0.8	1.1
100	AR	0.062	0.076	0.135	0.274	NA
	AR_{corr}	0.06	0.052	0.068	0.095	NA
	AR_{RLM}	0.063	0.061	0.053	0.071	0.069
	AR_{RL}	0.059	0.056	0.046	0.056	0.062
200	AR	0.062	0.088	0.145	0.279	NA
	AR_{corr}	0.059	0.066	0.058	0.079	NA
	AR_{RLM}	0.053	0.057	0.056	0.076	0.078
	AR_{RL}	0.051	0.054	0.043	0.05	0.045
500	AR	0.047	0.071	0.115	0.25	NA
	AR_{corr}	0.042	0.04	0.05	0.077	NA
	AR_{RLM}	0.055	0.062	0.073	0.098	0.128
	AR_{RL}	0.047	0.048	0.046	0.05	0.036

NB: AR, AR_{corr}, AR_{RL} and AR_{RLM} denote the conventional AR test, the modified AR test proposed in Anatolyev and Gospodinov (2011), the regularized AR Landweber-Frifman test proposed in this paper using bootstrap, and the regularized AR Landweber-Frifman test using simulated critical values, respectively.

Figure 1: Power curves of AR tests



on the fixed regularization parameter.

Tables 2 and 3 have the same results for PC and LF AR tests. The regularized AR tests using the bootstrap method lead in terms of overall performance.

To summarize, results for the AR tests show that the regularized AR test performs well relative to other AR tests. Indeed, the corrected Anatolyev and Gospodinov (2011)'s AR test allows us to correct the AR test size distortions but slightly over-rejects for large values of λ. The regularized AR has almost correct size results even with very large number of instruments (λ close to 1).

The Figure 1 gives the power curves of the AR_{RT}, AR_{RP} and AR_{RL}. The aim is to compare these tests as they are all sized correct. Based on Figure 1, we can conclude that the AR_{RT} statistic has typically better power properties than others regularized statistics.

In conclusion, these experiments suggest that using regularization as a dimension reduction tool can increase the power, while controlling for the size. Thus, regularization provides a promising solution to the size-power trade-offs arising from the use of many instruments, even in small samples.

5 Empirical application: Elasticity of Intertemporal Substitution

In a recent paper, Carrasco and Tchuente (2015) estimate EIS using many instrument and regularization. They follow Yogo (2004) who analyzes the problem of the estimation of the EIS using the linearized Euler equation. He explains how weak instruments have been the cause of the EIS empirical puzzle. He shows that, using conventional IV methods, the estimated EIS is significantly less than 1 but its reciprocal is not different from 1. Carrasco and Tchuente (2015) increase the number of instruments from 4 to 18 by including[5] interactions and power functions. Using the new instrument set, they propose a regularized 2SLS and LIML estimate of the EIS. The point estimates obtained by T and LF regularized estimators are very close to each other and are similar to those used for macro calibrations. However, we cannot reject the null hypothesis $H_0 : 1/\psi = 1$ or $H_0 : \psi = 1$ using t-test. This paper proposed confidence intervals for $1/\psi$ and $\psi = 1$ using regularized AR tests.

In this section, we follow the specifications in Yogo (2004) using quarterly data from 1947.3 to 1998.4 for the United States. The estimated model is given by:

$$\Delta c_{t+1} = \tau + \psi r_{f,t+1} + \xi_{t+1}$$

and the "reverse regression":

$$r_{f,t+1} = \mu + \frac{1}{\psi}\Delta c_{t+1} + \eta_{t+1}$$

where ψ is the EIS, Δc_{t+1} is the consumption growth at time $t+1$, $r_{f,t+1}$ is the real return on a risk free asset, μ and τ are constant, and η_{t+1} and ξ_{t+1} are respectively the shocks to asset return and to consumption.

Table 4 reports the 95% confidence intervals for the EIS constructed from the different AR tests. The Tikhonov (ridge) regularized AR test give confidence intervals that are small and consistent for ψ and $1/\psi$. The confidence interval for EIS $(1/\psi)$ is $[0.252, 0.71]$ for the Tikhonov bootstrap based test and $[0.26, 0.741]$ for the simulation based tests using both 4 and 18 instruments. The second regularized AR test is based on the principal components

[5] In his paper Yogo (2004) uses four instruments: the twice lagged, nominal interest rate (r), inflation (i), consumption growth (c) and log dividend-price ratio (p). This set of instruments is denoted $Z = [r, i, c, p]$. The 18 instruments used in our regression are derived from Z and are given by $II = [Z, Z.^2, Z.^3, Z(:,1) * Z(:,2), Z(:,1) * Z(:,3), Z(:,1) * Z(:,4), Z(:,2) * Z(:,3), Z(:,2) * Z(:,4), Z(:,3) * Z(:,4)]$, $Z.^k = [Z_{ij}^k]$, $Z(:,k)$ is the k^{th} column of Z and $Z(:,k) * Z(:,l)$ is a vector of interactions between columns k and l.

Table 4: Regularized AR test confidence interval for EIS.

	AR	AR_{corr}	AR_{RTM}	AR_{RT}	AR_{RPM}	AR_{RP}	AR_{RLM}	AR_{RL}
L=4								
ψ	$\emptyset$	$\emptyset$	[1.42, 4.1]	[1.42, 4.22]	[1.42, 4.1]	[1.42, 4.22]	[1.42, 4.1]	$[-\infty, +\infty]$
$1/\psi$	$\emptyset$	$\emptyset$	[0.23, 0.71]	[0.23, 0.71]	[0.23, 0.71]	[0.2, 0.67]	[0.23, 0.71]	$[-\infty, +\infty]$
L=18								
ψ	$\emptyset$	$\emptyset$	[1.42, 4.1]	[1.42, 4.22]	[1.42, 4.1]	[1.42, 4.22]	[1.42, 4.1]	$[-\infty, +\infty]$
$1/\psi$	$\emptyset$	$\emptyset$	[0.252, 0.71]	[0.26, 0.741]	[0.26, 0.741]	[0.26, 0.73]	[0.26, 0.731]	$[-\infty, +\infty]$

The table reports 95% confidence intervals for the EIS, constructed from AR,AR_{corr}, $AR_{R.}$ and $AR_{R.M}$ which denote respectively the conventional AR test, the modified AR test proposed in Anatolyev and Gospodinov (2011), the regularized AR test proposed in this paper using bootstrap, and the regularized AR test using simulated critical values (T, P and L are respectively for Tikhonov, Principal component and Landweber-Frifman). $\emptyset$ indicates an empty confidence interval. The instruments are twice lagged nominal interest rate, inflation, consumption growth, and log dividend-price ratio.

associated with the largest eigenvalues. This test confidence intervals for ψ are similar to those of the regularized Tikhonov AR. The last AR test is based on an iterative method called Landweber-Fridman. The regularized LF AR test using the bootstrap method suffers from identification failure as evidenced by the uninformative confidence intervals $[-\infty, +\infty]$ for both the $L = 4$ and $L = 18$. However, its simulated test based analogue gives the same conclusion with other tests.

As expected from Yogo (2004), the classic AR test and the corrected Anatolyev and Gospodinov (2011) AR lead to empty confidence sets for EIS.

In summary, the regularized AR confidence intervals indicate that the EIS is less than one and is in the range of the parameters used in the literature for macro calibrations.

6 Conclusion

This paper introduces the use of regularization techniques for inference. We investigate the use of three regularization techniques to construct robust to weak instruments procedures when the number of instrument is large. We propose three regularized AR tests. Inference can be performed using a new restricted efficient bootstrap method or simulated Monte Carlo test. Our simulations suggest that the regularized AR tests perform well and have better performance than Anatolyev and Gospodinov (2011) corrected AR test when the number of instruments is very large. An empirical application, of the estimation of the EIS suggests that the value of the the EIS is in the range of (0.2 , 0.76). Another topic of interest is the use of regularization to provide versions of robust test for weak instruments as Lagrange Multiplier (LM) or conditional likelihood ratio test (CLR) tests, that can be used with large numbers or a continuum of moments conditions.

References

Anatolyev, S., and N. Gospodinov (2011): "Specification Testing In Models With Many Instruments," *Econometric Theory*, 27(02), 427–441.

Anderson, T. W., and H. Rubin (1949): "Estimation of the parameters of a single equation in a complete system of stochastic equations," *The Annals of Mathematical Statistics*, pp. 46–63.

Andrews, D. W., and J. H. Stock (2007): "Testing with many weak instruments," *Journal of Econometrics*, 138(1), 24–46.

Angrist, J. D., and A. B. Krueger (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?," *The Quarterly Journal of Economics*, 106(4), 979–1014.

Arellano, M., and S. Bond (1991): "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *The review of economic studies*, 58(2), 277–297.

Bekker, P. A. (1994): "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, 62(3), 657–81.

Carrasco, M. (2012): "A regularization approach to the many instruments problem," *Journal of Econometrics*, 170(2), 383–398.

Carrasco, M., and J.-P. Florens (2000): "Generalization Of Gmm To A Continuum Of Moment Conditions," *Econometric Theory*, 16(06), 797–834.

——— (2014): "On the asymptotic efficiency of GMM," *Econometric Theory*, 30(02), 372–406.

Carrasco, M., J.-P. Florens, and E. Renault (2007): "Linear Inverse Problems in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6 of *Handbook of Econometrics*, chap. 77. Elsevier.

Carrasco, M., and B. Rossi (2016): "In-sample Inference and Forecasting in Misspecified Factor Models," *Journal of Business & Economic Statistics*, (just-accepted), 1–72.

Carrasco, M., and G. Tchuente (2015): "Regularized LIML for many instruments," *Journal of Econometrics*, 186(2), 427–442.

Chao, J. C., and N. R. Swanson (2005): "Consistent Estimation with a Large Number of Weak Instruments," *Econometrica*, 73(5), 1673–1692.

Davidson, R., and J. G. MacKinnon (1999): "Bootstrap Testing in Nonlinear Models," *International Economic Review*, 40(2), 487–508.

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

Davidson, R., and J. G. MacKinnon (2014a): "Bootstrap confidence sets with weak instruments," *Econometric Reviews*, 33(5-6), 651–675.

——— (2014b): "Confidence sets based on inverting Anderson–Rubin tests," *The Econometrics Journal*, 17(2), S39–S58.

Dufour, J.-M. (1997): "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 65(6), 1365–1388.

Dufour, J.-M., and M. Taamouti (2007): "Further results on projection-based inference in IV regressions with weak, collinear or missing instruments," *Journal of Econometrics*, 139(1), 133–153.

Dufour, J.-M., and P. Valéry (2016): "Wald-type tests when rank conditions fail: a smooth regularization approach," Discussion paper.

Flores-Lagunes, A. (2007): "Finite sample evidence of IV estimators under weak instruments," *Journal of Applied Econometrics*, 22(3), 677–694.

Hall, P., and J. L. Horowitz (1996): "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators," *Econometrica*, 64(4), 891–916.

Hansen, C., J. Hausman, and W. Newey (2008): "Estimation With Many Instrumental Variables," *Journal of Business & Economic Statistics*, 26, 398–422.

Kapetanios, G., L. Khalaf, and M. Marcellino (2015): "Factor-Based Identification-Robust Interference in IV Regressions," *Journal of Applied Econometrics*, pp. n/a–n/a.

Kleibergen, F. (2002): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70(5), 1781–1803.

Leucht, A., and M. H. Neumann (2009): "Consistency of general bootstrap methods for degenerate U- and V -type statistics.," *J. Multivariate Anal.*, 100, 1622–1633.

——— (2012): "Dependent wild bootstrap for degenerate U- and V -statistics," Discussion paper.

Moreira, M. J. (2003): "A conditional likelihood ratio test for structural models," *Econometrica*, 71(4), 1027–1048.

Newey, W. K., and F. Windmeijer (2009): "Generalized Method of Moments With Many Weak Moment Conditions," *Econometrica*, 77(3), 687–719.

Serfling, R. (1980): *Approximation Theorems of Mathematical Statistics*. New York: Wiley.

Wang, W., and M. Kaffo (2014): "Bootstrap Inference for Instrumental Variable Models with Many Weak Instruments," Working paper, University of Montreal and Kyoto University.

Yogo, M. (2004): "Estimating the Elasticity of Intertemporal Substitution When Instruments Are Weak," *The Review of Economics and Statistics*, 86(3), 797–810.

A Regularized methods

This section present the regularization methods. These methods are the same as those used in Carrasco (2012). We use a compact notation which allows us to deal with a finite, countable infinite number of moments, or a continuum of moments.

Let us consider the following sequence of instruments, $Z_i = Z(\tau; x_i)$ where $\tau \in S$ may be an integer or an index taking its values in an interval. Examples of Z_i are:

- $Z_i = x_i$ where x_i is a L- vector with a fixed L. Then $Z(\tau; x_i)$ denotes the τth element of x_i and $S = \{1, 2, ...L\}$.
- $Z_{ij} = (x_i)^{j-1}$ with $j \in S = \mathbb{N}$, thus we have an infinite countable instruments.
- $Z_i = Z(\tau; x_i) = exp(i\tau' x_i)$ where $\tau \in S = \mathbb{R}^{dim(x_i)}$, thus we have a continuum of moments.

We now define the covariance operator for the instrumental variable set. We denote $L^2(\pi)$ the Hilbert space of square integrable functions with respect to π where π is a positive measure on S. For a detailed discussion on the role of π, see Carrasco (2012).

The covariance operator K of the instruments is

$$\mathbf{K} : L^2(\pi) \rightarrow L^2(\pi)$$
$$(\mathbf{K}g)(\tau_1) = \int E[Z(\tau_1; x_i)\overline{Z(\tau_2; x_i)}]g(\tau_2)\pi(\tau_2)d\tau_2$$

where $\overline{Z(\tau_2; x_i)}$ denotes the complex conjugate of $Z(\tau_2; x_i)$.

$\mathbf{K}$ is assumed to be a compact operator (see Carrasco, Florens, and Renault (2007) for a definition). Carrasco and Florens (2014) show that π can be chosen so that $\mathbf{K}$ is compact.

Let λ_j and ϕ_j $j = 1, 2, ...$ be respectively the eigenvalues (ordered in decreasing order) and the orthogonal eigenfunctions of $\mathbf{K}$. The operator $\mathbf{K}$ can be estimated by $\mathbf{K}_n$ defined as:

$$\mathbf{K}_n : L^2(\pi) \rightarrow L^2(\pi)$$
$$(\mathbf{K}_n g)(\tau_1) = \int \frac{1}{n}\sum_{i=1}^{n} Z(\tau_1; x_i)\overline{Z(\tau_2; x_i)}g(\tau_2)\pi(\tau_2)d\tau_2$$

If the number of moment conditions is infinite, the inverse of $\mathbf{K}_n$ needs to be regularized because it is nearly singular. By definition (see Kress, 1999, page 269), a regularized inverse of an operator $\mathbf{K}$ is

$$R_\alpha : L^2(\pi) \rightarrow L^2(\pi)$$

such that $\lim_{\alpha \to 0} R_\alpha \mathbf{K}\varphi = \varphi$, $\forall \varphi \in L^2(\pi)$.

We consider three different types of regularization schemes: Tikhonov (T); Landwerber Fridman (LF); Spectral cut-off (SC). They are defined as follows:

1. **Tikhonov(T)**

   This regularization scheme is closely related to the ridge regression[6].

   $(\mathbf{K}^{\alpha})^{-1} = (\mathbf{K}^2 + \alpha I)^{-1}\mathbf{K}$

   $$(\mathbf{K}^{\alpha})^{-1} r = \sum_{j=1}^{\infty} \frac{\lambda_j}{\lambda_j^2 + \alpha} \langle r, \phi_j \rangle \phi_j$$

   where $\alpha > 0$ and I is the identity operator.

2. **Landweber-Fridman (LF)**

   This method of regularization is iterative. Let $0 < c < 1/\|\mathbf{K}\|^2$ where $\|\mathbf{K}\|$ is the largest eigenvalue of $\mathbf{K}$ (which can be estimated by the largest eigenvalue of $\mathbf{K}_n$). $\hat{\varphi} = (\mathbf{K}^{\alpha})^{-1} r$ is computed using the following procedure:

   $$\begin{cases} \hat{\varphi}_l = (1 - c\mathbf{K}^2)\hat{\varphi}_{l-1} + c\mathbf{K}r, \quad \text{l=1,2,...,}\dfrac{1}{\alpha} - 1; \\ \hat{\varphi}_0 = c\mathbf{K}r, \end{cases}$$

   where $\frac{1}{\alpha} - 1$ is some positive integer. We also have

   $$(\mathbf{K}^{\alpha})^{-1} r = \sum_{j=1}^{\infty} \frac{[1 - (1 - c\lambda_j^2)^{\frac{1}{\alpha}}]}{\lambda_j} \langle r, \phi_j \rangle \phi_j.$$

3. **Spectral cut-off (SC)**

   This method consists in selecting the eigenfunctions associated with the eigenvalues greater than some threshold. The aim is to select those who have greater contribution.

   $$(\mathbf{K}^{\alpha})^{-1} r = \sum_{\lambda_j^2 \geq \alpha} \frac{1}{\lambda_j} \langle r, \phi_j \rangle \phi_j$$

   for $\alpha > 0$.

   This method is similar to principal components (PC) which consists in using the first eigenfunctions:

   $$(\mathbf{K}^{\alpha})^{-1} r = \sum_{j=1}^{1/\alpha} \frac{1}{\lambda_j} \langle r, \phi_j \rangle \phi_j$$

   where $\frac{1}{\alpha}$ is some positive integer.

[6] $\langle .,. \rangle$ represents the scalar product in $L^2(\pi)$ and in $\mathbb{R}^n$ (depending on the context).

These regularized inverses can be rewritten in common notation as:

$$(\mathbf{K}^{\alpha})^{-1}r = \sum_{j=1}^{\infty}\frac{q(\alpha,\lambda_j^2)}{\lambda_j}\langle r,\phi_j\rangle\phi_j$$

where for T: $q(\alpha,\lambda_j^2) = \dfrac{\lambda_j^2}{\lambda_j^2+\alpha}$,
for LF: $q(\alpha,\lambda_j^2) = [1-(1-c\lambda_j^2)^{1/\alpha}]$,
for SC: $q(\alpha,\lambda_j^2) = I(\lambda_j^2 \geq \alpha)$, for PC $q(\alpha,\lambda_j^2) = I(j \leq 1/\alpha)$.

In order to compute the inverse of $\mathbf{K}_n$ we have to choose the regularization parameter α.

$(\mathbf{K}_n^{\alpha})^{-1}$ is the regularized inverse of $\mathbf{K}_n$ and P^{α} a $n \times n$ matrix defined as in Carrasco (2012) by $P^{\alpha} = T(\mathbf{K}_n^{\alpha})^{-1}T^*$ where

$$T : L^2(\pi) \rightarrow \mathbb{R}^n$$

$$Tg = \begin{pmatrix} \langle Z_1, g\rangle \\ \langle Z_2, g\rangle \\ . \\ . \\ \langle Z_n, g\rangle \end{pmatrix}$$

and

$$T^* : \mathbb{R}^n \rightarrow L^2(\pi)$$

$$T^*v = \frac{1}{n}\sum_{i=1}^{n} Z_i v_i$$

such that $K_n = T^*T$ and TT^* is an $n \times n$ matrix with typical element $\dfrac{\langle Z_i, Z_j\rangle}{n}$. Let $\hat{\phi}_j$, $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq ... > 0$, $j = 1,2,...$ be the orthonormalized eigenfunctions and eigenvalues of K_n. $\hat{\lambda}_j$ are consistent estimators of λ_j, the eigenvalues of TT^*. We then have $T\hat{\phi}_j = \sqrt{\lambda_j}\psi_j$ and $T^*\psi_j = \sqrt{\lambda_j}\hat{\phi}_j$.

For $v \in \mathbf{R}^n$, $P^{\alpha}v = \sum_{j=1}^{\infty} q(\alpha,\lambda_j^2)\langle v,\psi_j\rangle\psi_j$. It follows that for any vectors v and w of $\mathbf{R}^n$:

$$\begin{aligned} v'P^{\alpha}w &= v'T(K_n^{\alpha})^{-1}T^*w \\ &= \left\langle (K_n^{\alpha})^{-1/2}\sum_{i=1}^{n} Z_i\,(.)\,v_i, (K_n^{\alpha})^{-1/2}\frac{1}{n}\sum_{i=1}^{n} Z_i\,(.)\,w_i \right\rangle. \end{aligned} \quad (7)$$

P^{α} is the regularized version of the projection matrix on the space spanned by the set of instruments.

B Proofs

Proof of Theorem 1:

This section presents the derivation of the asymptotic behavior of the $AR_R(\delta_0)$ statistic. We know that

$$AR_R(\delta_0) = \frac{n(y - W\delta_0)' P^\alpha (y - W\delta_0)}{(y - W\delta_0)'[I_n - P^\alpha](y - W\delta_0)}$$

and $\varepsilon = (y - W\delta_0)$. Hence,

$$\begin{aligned} AR_R(\delta_0) &= \frac{n\varepsilon' P^\alpha \varepsilon}{\varepsilon'[I_n - P^\alpha]\varepsilon} \\ &= \frac{\varepsilon' P^\alpha \varepsilon}{\sigma_\varepsilon^2} \frac{n\sigma_\varepsilon^2}{\varepsilon'[I_n - P^\alpha]\varepsilon} \\ &= J_0 \left[\frac{\varepsilon'\varepsilon}{n\sigma_\varepsilon^2} - J_0/n \right]^{-1} \end{aligned}$$

where $J_0 = \dfrac{\varepsilon' P^\alpha \varepsilon}{\sigma_\varepsilon^2}$.

From lemma 1 Carrasco (2012), $J_0 = O_p(\frac{1}{\alpha})$ and by Taylor expansion.

$$\begin{aligned} AR_R(\delta_0) &= J_0 \left[1 - (\frac{\varepsilon'\varepsilon}{n\sigma_\varepsilon^2} - 1) + O_p(\frac{1}{n\alpha}) \right]^{-1} \\ &= J_0 + \left(1 - \frac{\varepsilon'\varepsilon}{n\sigma_\varepsilon^2} \right) + O_p(\frac{1}{n\alpha}) \\ &= J_0 + O_p(\frac{1}{n\alpha}) \end{aligned}$$

Thus, up to an $O_p\left(\dfrac{1}{n\alpha}\right)$ remainder,

$$\begin{aligned} AR_R(\delta_0) &= J_0 \\ &= \sum_{i=1}^n P_{ii}^\alpha \frac{\varepsilon_i^2}{\sigma_\varepsilon^2} + \frac{1}{n} \sum_{i \neq j} P_{ij}^\alpha \frac{\varepsilon_i \varepsilon_j}{\sigma_\varepsilon^2} \end{aligned}$$

We can derived the asymptotic behavior of regularized AR test using two methods. We first use results on V-statistics and after that we will derive the same asymptotic by definition of P^α and asymptotic behavior established in Carrasco and Florens (2000).

Let us define $\tilde{x}_i = \dfrac{Z_i \varepsilon_i}{\sigma_\varepsilon}$ and $h(\tilde{x}_i, \tilde{x}_j) = \left\langle \tilde{x}_i, (K^\alpha)^{-1} \tilde{x}_j \right\rangle$ (remember that $\tilde{x}_i$ is a function indexed by τ because Z_i is also a function of τ, such a representation can handle both countable and continuum of instruments).

We can write J_0 as a degenerated V-statistics[7]. Using definition of P^α in equation (7).

$$J_0 = V_n = \frac{1}{n}\sum_{i,j} h(\tilde{x}_i, \tilde{x}_j)$$

Where $h(x_1, x_2) = \sum_{l=1}^{\infty} q_l \Phi_l(x_1)\Phi_l(x_2)$ with $\Phi_l(x_1) = \langle x_1, \psi_j \rangle$.

We now prove that

$$AR_R(\delta_0) \to^d \sum_{l=1}^{\infty} q_l \chi_l^2(1)$$

Let us replace h by its decomposed value in V_n,

$$V_n = \frac{1}{n}\sum_{i,j}\sum_{l=1}^{\infty} q_l \Phi_l(\tilde{x}_i)\Phi_l(\tilde{x}_j)$$

And let us define $V_{n,M} = \frac{1}{n}\sum_{i,j}\sum_{e=1}^{M} q_e \Phi_e(\tilde{x}_i)\Phi_e(\tilde{x}_j)$,

$V = \sum_{l=1}^{\infty} q_l \chi_l^2(1)$,

and $V_M = \sum_{e=1}^{M} q_e \chi_e^2(1)$.

To derive the asymptotic distribution of $AR_R(\delta_0)$, we show that $Ee^{itV_n} \to Ee^{itV}$.

$$\begin{aligned}
|Ee^{itV_n} - Ee^{itV_{nM}}| &\leq E|e^{itV_n} - e^{itV_{nM}}| \\
&\leq E|e^{it(V_n - V_{nM})} - 1| \\
&\leq |t|E|V_n - V_{nM}| \\
&\leq |t|\left[E(V_n - V_{nM})^2\right]^{\frac{1}{2}}
\end{aligned}$$

The second inequality uses $|e^{iz} - 1| \leq |z|$. Using the definition of V_n and V_{nM} we have

$$E(V_n - V_{nM})^2 \leq c_\alpha (\sum_{e=M+1}^{\infty} q_e^2)^2$$

Let t be fixed and ϵ be given.

Therefore, for M large enough, under Assumption 2,

$$|t|(c_\alpha \sum_{e=M+1}^{\infty} q_e^2) \leq \epsilon$$

[7] See Leucht and Neumann (2009) and Leucht and Neumann (2012) for some information on V-statistics

This establishes that for M large enough and all n

$$|Ee^{itV_n} - Ee^{itV_{nM}}| \leq \epsilon$$

Now let us show that $V_{nM} \xrightarrow{d} V_M$

$h_n = \dfrac{1}{\sqrt{n}} \sum_{1}^{n} \dfrac{Z_i \varepsilon_i}{\sigma_\varepsilon}$ and $W_{ne} = \langle h_n, \phi_e \rangle / \sqrt{\lambda_e}$
where $\{\phi_e\}_{e=1}^{\infty}$ are orthonormal eigenfunctions.

The compactness of $\mathbf{K}$ implies a functional central limit theorem, h_n converges in $L^2(\pi)$ to a mean zero Gaussian process with covariance operator $\sigma_\varepsilon^2 K$.
Thus, $W_{ne} \xrightarrow{d} \mathcal{N}(0,1)$ and are all independent.
We can then conclude for n sufficiently large that

$$|Ee^{itV_{nM}} - Ee^{itV_M}| \leq \epsilon.$$

Thus, $V_{nM} \xrightarrow{d} V_M$

By definition of V and V_M, we show that for M large enough

$$|Ee^{itV} - Ee^{itV_M}| \leq \epsilon$$

And we can conclude that $|Ee^{itV_n} - Ee^{itV}| \leq 3\epsilon$, for any t and any ϵ, and all n sufficiently large.

It follows that

$$AR_R(\delta_0) \xrightarrow{d} \sum_{l=1}^{\infty} q_l \chi_l^2(1)$$

This prove uses the same argument as in Serfling (1980)(pp. 193-194).

Now let us establish the same results using only definition of P^α.

$$AR_R(\delta_0) = J_0 + O_p\left(\frac{1}{n\alpha}\right)$$

with $J_0 = \dfrac{\varepsilon' P^\alpha \varepsilon}{\sigma_\varepsilon^2}$ using definition of P^α and h_n one can show that

$$\begin{aligned} J_0 &= \sum_{l=1}^{\infty} \frac{q_l}{\lambda_l} \langle h_n, \phi_l \rangle^2 \\ &= \sum_{l=1}^{\infty} q_l \frac{\langle h_n, \phi_l \rangle^2}{\lambda_l} \end{aligned}$$

Following Carrasco and Florens (2000) we can show that $\frac{\langle h_n, \phi_l \rangle^2}{\lambda_l} \xrightarrow{d} \chi^2(1)$ and are all independently distributed.

It follows that

$$AR_R(\delta_0) \xrightarrow{d} \sum_{l=1}^{\infty} q_l \chi_l^2(1)$$

To end the proof we discuss how $\hat{\lambda}_j \xrightarrow{p} \lambda_j$ affects the asymptotic value of the regularized AR test. $\hat{\lambda}_j$ is the a consistent estimate of the j^{th} eigenvalue. For any α, q_j is a continuous function of λ_j and $\sum_{l=1}^{\infty} q_l \chi_l^2(1) < \infty$ is a continuous function of all λ_j. Using the continuous mapping theorem, we can conclude that

$$\sum_{l=1}^{\infty} \hat{q}_l \chi_l^2(1) \xrightarrow{d} \sum_{l=1}^{\infty} q_l \chi_l^2(1)$$

This ends the proof of Theorem 1.

Proof of Theorem 2:

Here we show that the $AR_R^*(\delta_0)$ statistic that we define have the same asymptotic distribution in the bootstrap world as the original one.

$$\begin{aligned} AR_R^*(\delta_0) &= \frac{n\varepsilon^{*\prime} P^{\alpha} \varepsilon *}{\varepsilon^{*\prime}[I_n - P^{\alpha}]\varepsilon^*} \\ &= \frac{\varepsilon^{*\prime} P^{\alpha} \varepsilon^*}{\sigma_{\varepsilon^*}^2} \frac{\sigma_{\varepsilon^*}^2}{\frac{1}{n}\varepsilon^{*\prime}[I_n - P^{\alpha}]\varepsilon^{*\prime}} \\ &= J_0^* \left[\frac{\varepsilon^{*\prime}\varepsilon^*}{n\sigma_{\varepsilon^*}^2} - \frac{J_0^*}{n} \right]^{-1} \end{aligned}$$

where $J_0^* = \frac{\varepsilon^{*\prime} P^{\alpha} \varepsilon^*}{\sigma_{\varepsilon^*}^2}$ with $\sigma_{\varepsilon^*}^2 = E^*(\varepsilon_i^* \varepsilon_i^*)$

Let us first show that $E^*(\varepsilon_i^* \varepsilon_i^*)$ is bounded under Assumption 3.

Let us define $\bar{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i$, $\bar{W} = \frac{1}{n}\sum_{i=1}^{n} W_i$

Using Minkowski and Cauchy-Schwartz inequalities, we show that

$$\begin{aligned}
E^*(\varepsilon_i^*\varepsilon_i^*) &= \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^*\varepsilon_i^* \\
&= \frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i - \bar{\varepsilon} - (W_i - \bar{W})'(\hat{\delta} - \delta))^2 \\
&\leq c_1\left[\frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i - \bar{\varepsilon})^2 + \frac{1}{n}\sum_{i=1}^{n}((W_i - \bar{W})'(\hat{\delta} - \delta))^2\right] \\
&\leq c_2\left[\frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i - \bar{\varepsilon})^2 + \|\hat{\delta} - \delta)\|^2\frac{1}{n}\sum_{i=1}^{n}\|W_i - \bar{W}\|^2\right] \\
&\leq c_3\left[\frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i^2 - \bar{\varepsilon}^2 + \|\hat{\delta} - \delta)\|^2\frac{1}{n}\sum_{i=1}^{n}\|W_i\|^2 + \|\bar{W}\|^2\right].
\end{aligned}$$

Under Assumption 3 we can then conclude that $E^*(\varepsilon_i^*\varepsilon_i^*)$ is bounded.

It is therefore possible to show that

$$E^*[(J_0^*)/n] = \frac{tr(P^\alpha E^*((\varepsilon^{*'}\varepsilon^*)|X)}{n\sigma^2_{\varepsilon^{*'}}} = O_{p^*}(\frac{1}{n\alpha})$$

by the Markov inequality.

Next we can show that

$$\begin{aligned}
AR_R^*(\delta_0) &= J_0^*\left[1 - (\frac{\varepsilon^{*'}\varepsilon^*}{n\sigma^2_{\varepsilon^*}} - \frac{J_0^*}{n} - 1) + O_{p^*}(\frac{1}{n\alpha})\right] \\
&= J_0^* + O_{p^*}(\frac{1}{n\alpha})
\end{aligned}$$

Thus, up to an $O_{p^*}(\frac{1}{n\alpha})$ remainder,

$$\begin{aligned}
AR_R^*(\delta_0) &= J_0^* \\
&= \sum_{i=1}^{n} P_{ii}^\alpha \frac{\varepsilon_i^{*2}}{\sigma^2_{\varepsilon_i^*}} + \sum_{i\neq j} P_{ij}^\alpha \frac{\varepsilon_i^*\varepsilon_j^*}{\sigma^2_{\varepsilon_i^*}}
\end{aligned}$$

We have showed that the $AR_R^*(\delta_0)$ are up to negligible remainder terms V-statistics. Therefore, following the same arguments as in Proposition 1, but now in the bootstrap world, we show that $AR_R^*(\delta_0) \xrightarrow{d^*} \sum_{k=1}^{\infty} q_l\chi_l^2(1)$.

And the second result follows by Polya's Theorem, because the $\sum_{k=1}^{\infty} q_l\chi_l^2(1)$ distribution is everywhere continuous.

Recent Kent Discussion Papers in Economics

16/07: 'Estimation of social interaction models using regularization', Guy Tchuente

16/06: 'The Post-crisis Slump in Europe: A Business Cycle Accounting Analysis', Florian Gerth and Keisuke Otsu

16/05: 'The Revenue Implication of Trade Liberalisation in Sub-Saharan Africa: Some new evidence', Lanre Kassim

16/04: 'The rise of the service economy and the real return on capital', Miguel León-Ledesma and Alessio Moro

16/03: 'Is there a mission drift in microfinance? Some new empirical evidence from Uganda', Francis Awuku Darko

16/02: 'Early Marriage, Social Networks and the Transmission of Norms', Niaz Asadullah and Zaki Wahhaj

16/01: 'Intra-household Resource Allocation and Familial Ties', Harounan Kazianga and Zaki Wahhaj

15/21: 'Endogenous divorce and human capital production', Amanda Gosling and María D. C. García-Alonso

15/20: 'A Theory of Child Marriage', Zaki Wahhaj

15/19: 'A fast algorithm for finding the confidence set of large collections of models', Sylvain Barde

15/18: 'Trend Dominance in Macroeconomic Fluctuations', Katsuyuki Shibayama

15/17: 'Efficient estimation with many weak instruments using regularization techniques', Marine Carrasco and Guy Tchuente

15/16: 'High school human capital portfolio and college outcomes', Guy Tchuente

15/15: 'Regularized LIML for many instruments, Marine Carrasco and Guy Tchuente

15/14: 'Agglomeration Economies and Productivity Growth: U.S. Cities, 1880-1930', Alexander Klein and Nicholas Crafts

15/13: 'Microcredit with Voluntary Contributions and Zero Interest Rate - Evidence from Pakistan', Mahreen Mahmud

15/12: 'Act Now: The Effects of the 2008 Spanish Disability Reform', Matthew J. Hill, Jose Silva and Judit Vall

15/11: 'Testing for Level Shifts in Fractionally Integrated Processes: a State Space Approach', Davide Delle Monache, Stefano Grassi and Paolo Santucci de Magistris