FEDERAL RESERVE BANK *of* ATLANTA **WORKING PAPER SERIES**

A Uniformly Valid Test for Instrument Exogeneity

Prosper Dovonon and Nikolay Gospodinov

Working Paper 2025-9
September 2025

Abstract: This paper studies the limiting behavior of the test for instrument exogeneity in linear models when there is uncertainty about the strength of the identification signal. We consider the test for conditional moment restrictions with an expanding set of constructed instruments. We establish the uniform validity of the standard normal asymptotic approximation, under the null, of this specification test over all possible degrees of model identification. As a result, this allows the researcher to use standard inference for testing instrument exogeneity without the need of any prior knowledge if the instruments are strong, semi-strong, weak, or completely irrelevant. Furthermore, we show that the test is consistent regardless of the instrument strength; i.e., even in cases (weak and completely irrelevant instruments) where the standard tests fail to exhibit asymptotic power. To obtain these results, we characterize the rate of the estimator under a drifting sequence for the identification signal. We illustrate the appealing properties of the test in simulations and an empirical application.

JEL classification: C12, C14, C26, C52

Key words: linear instrumental variables (IV) model, conditional test for instrument exogeneity, uniform inference, instrument strength, generalized method of moments (GMM) estimator, drifting sequences, expanding set of basis functions

https://doi.org/10.29338/wp2025-09

Prosper Dovonon acknowledges financial support from the Social Sciences and Humanities Research Council of Canada. The views expressed here are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

Prosper Dovonon (corresponding author) is in the Economics Department, Concordia University, 1455 de Maisonneuve Blvd. West, H 1155, Montreal, Quebec, Canada, H3G 1M8, prosper.dovonon@concordia.ca. Nikolay Gospodinov is in the Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree Street NE, Atlanta, GA 30309-4470, nikolay.gospodinov@atl.frb.org.

1 Introduction

Standard inference in linear instrumental variable (IV) models depends critically on two properties of the instruments: (i) their strength or relevance in explaining the variation in the endogenous variables and (ii) their exogeneity that ensures correct specification of the model moment conditions. Given that models should always be viewed only as approximations of a possibly unknowable data generating process, it is often prudent to assess the degree of model misspecification by subjecting the moment conditions to a test for correct specification by remaining agnostic about the strength of the identification signal of the instruments. Unfortunately, standard specification tests in the existing literature are not robust to uncertainty about the instrument strength and their limiting behavior (under the null and the alternative) tends to break down when the instruments are weak or irrelevant. For example, under failure of the rank condition for identification, the conventional tests for the validity of the overidentifying restrictions in linear models have a non-standard limit under the null and are inconsistent under the alternative hypothesis (Cragg and Donald, 1996; Kitamura, 2006; Gospodinov, Kan and Robotti, 2017; among others). Since the access to strong observable instruments that can point-identify the economic model is often constrained, detecting model misspecification – such as invalidity of the instruments – in an identification-robust way appears elusive. And yet, many interesting economic models of policy relevance are characterized by the presence of weak (or outright irrelevant) but possibly endogenous instruments that lead to the violation of the model moment conditions (see, for example, Murray, 2006; Bazzi and Clemens, 2013; among many others).

When the dimension of the instrument vector, k, is large relative to the sample size n, the tests for overidentifying restrictions can be modified to reflect the expanding set of instruments. More specifically, while the Sargan-Hansen J test for overidentifying restrictions has a chi-squared limit when k is fixed, its scaled and recentered version $S_{n,k} = (J_{n,k} - k)/\sqrt{2k}$ has a standard normal limit under the condition $k = o(n^{1/3})$; see Donald, Imbens and Newey (2003). Similar adjustments can render the Anderson-Rubin test asymptotically normal (Andrews and Stock, 2007) with moderately many instruments. In Bekker's (1994) many instrument framework, where the instruments grow at rate that is proportional to the sample size, Anatolyev and Gospodinov (2011) construct corrected versions of these tests that are robust to the numerosity of the instruments and are valid under both the fixed and many instrument asymptotics. However, these results require that the IV estimator is root-n consistent; a rate that may not be attainable when the degree of identification of the model is compromised.[1] Thus, it seems desirable to have a consistent test, with a standard limiting distribution under the null, that remains uniformly valid, irrespective of the identification strength of the instruments.

It is instructive, at this point, to compare and contrast the tests for unconditional moment restrictions, described above, to the tests for validity of conditional moment restrictions, developed by Bierens (1982), Bierens and Ploberger (1987), de Jong and Bierens (1994), Carrasco and Florens (2000), Donald, Imbens and Newey (2003), Tripathi and Kitamura (2003), among others. First, many economic models are defined by a set of conditional moment restrictions with a small number of conditioning variables. The expanding number of instruments (or continuum of moment conditions), k, is then constructed as a sequence of basis functions of the conditioning variables, which is completely under the control of the researcher. This approach has a substantial practical advantage because it does not require access to many economic variables that can serve as instruments.[2] While in the latter case the choice and justification of the relationship between k and n can be arbitrary, in the conditional moment restriction test, the expanding set of basis functions k can be obtained in practice precisely as the limiting theory (under the null or the alternative) requires. With this distinction in mind, the conditional specification test has also been shown to converge asymptotically to the standard normal distribution, provided that $k \to \infty$ at some rate and the estimator is root-n consistent; i.e., under the maintained assumption that the identification or rank condition is satisfied.

In this paper, we build on these strands of literature and establish the uniform validity of the test for instrument exogeneity in linear models which is completely agnostic to the degree of model identification, without the need of any prior knowledge of whether the instruments are strong, semi-strong, weak or completely irrelevant. The test statistic $S_{n,k}$, based on k basis functions $g(z_i)$ of some finite number of conditioning variables or potential instruments z_i, is pivotal under the null of instrument exogeneity $H_0 : \Pr(E(\varepsilon_i|z_i) = 0) = 1$, where ε_i denotes the regression errors. As a result, the test is straightforward to implement as it uses standard normal critical values with a tuning parameter (k) that is fully under the control of the researcher.

To characterize the full range of possibilities for the identification signal, we cast it as a drifting

[1] Under some conditions and parameterizations (see Chao and Swanson, 2005; Han and Phillips, 2006; Mikusheva and Sun, 2022), an expanding set of weak instruments may enhance the identification signal and render the estimator consistent.

[2] While there are situations in which a large number of instruments can be constructed by interacting different variables (Angrist and Krueger, 1991) or using lagged dependent variables in panel data models (Arellano and Bond, 1991), this is not always the case and invoking the many instrument asymptotics may be challenging.

parameter sequence of the sample size. More specifically, we parameterize the conditional mean of the endogenous variables x_i given z_i as $C(z_i)/n^\delta$ for some arbitrary, but not necessarily linear, measurable localizing function $C(z)$ and a scalar parameter $\delta \in [0, +\infty]$ that controls the degree of identification. To obtain the asymptotic behavior of the specification test S_n, we establish the limiting properties of the generalized method of moments (GMM) estimator over the range of values of δ under both the null and alternative hypotheses. First, for any $\delta \geq 1/2$, which represents "weak instrument" region and includes the case of completely irrelevant instruments ($\delta = +\infty$), we show that the GMM estimator has a probability limit but is inconsistent. Furthermore, for $0 \leq \delta < 1/2$ – which covers the strong instrument ($\delta = 0$) and semi strong/weak instrument case – consistent estimation is possible but it hinges on the choice k. If k grows at a slower rate than $n^{1-2\delta}$, then the GMM estimator is consistent. If k grows at the same rate as $n^{1-2\delta}$ or faster, then the consistency is lost but the convergence to a probability limit is preserved. We then use these results to establish the uniform validity, under the null, of the $N(0,1)$ limit of $S_{n,k}$ over $\delta \in [0, +\infty]$.

Under the alternative, $H_1 : \Pr\left(E(\varepsilon_i|z_i) = 0\right) < 1$, the limiting behavior of the GMM estimator is characterized by the interaction between the invalidity of the instruments and their identification strength. When the instruments are weak or completely irrelevant ($\delta \geq 1/2$), the GMM estimator diverges at rate $\sqrt{n}/k$ while the GMM estimator converges to its pseudo-true value when the instruments are strong ($\delta = 0$). These results, along with the limiting behavior of the optimal weighting matrix, ensure that the $S_{n,k}$ test statistic diverges to $+\infty$ under the alternative as the sample size grows. Interestingly, while the source of power is standard for $\delta \in [0, 1/2)$, in the case of weak or completely irrelevant instruments ($\delta \geq 1/2$) the power of the test is driven by the increasing number of generated instruments, k.

It may be beneficial to further position these results in the literature on specification testing with identification failure. Dovonon and Gospodinov (2024a) obtained conditions under which the specification test $S_{n,k}$ retains its standard normal limit when first-order local identification fails but global identification is still attainable. Doko Tchatoka and Dufour (2023) derive conditions for consistency of exogeneity tests in weakly identified IV models. The consistency of the tests, however, requires existence of at least some minimal signal in the instruments which is not satisfied in the case of completely irrelevant instruments which is covered by our theory (see, also, Caner, 2014).[3] Furthermore, in a setup where the conditional moment restrictions are estimated nonpara-

[3]When there is uncertainty about the strength of the identification signal, one could resort to identification-robust inference (see, for example, Kleibergen, 2005). However, Guggenberger (2012) shows that these methods do not

metrically, Jun and Pinkse (2009) obtain asymptotically valid specification tests without assuming identification. Their tests, however, are consistent only if identification is not too weak, which rules out the possibility of completely irrelevant instruments. Finally, Antoine and Lavergne (2023) also establish the uniform validity, irrespective of the identification strength, of a nonparametric (integrated conditional moment) statistic that tests jointly the value of the coefficient and the specification of the model (see also Stock and Wright, 2000). The asymptotic distribution of this test is non-pivotal and critical values are obtained by simulation. In contrast, our test is pivotal, easy to construct and implement, with uniform validity that is obtained in a richer, parameter-drifting setting. In closing, we should note that our arguments (under both the null and the alternative) would continue to go through, with minor modifications, if z_i itself is a high-dimensional vector such as in Kolesár *et al.* (2015), Guo *et al.* (2018), and Frandsen, Lefgren and Leslie (2023).

The rest of the paper is organized as follows. Section 2 introduces the conditional moment restriction setup and the main assumptions, the drifting parameterization of the identification signal and the expanding basis functions of the conditioning variables. Section 3 derives the asymptotic behavior of the estimator, weighting matrix and the tests statistic for instrument exogeneity under the null hypothesis. Section 4 characterizes the rates of the estimator and the weighting matrix under the alternative hypothesis and obtains the consistency of the specification test. Section 5 reports simulation results while Section 6 illustrates the practical benefits of the proposed test in an empirical application of the effect of international trade on economic growth. Section 7 concludes. The proofs of the main results are provided in Appendix A while some additional results are relegated to an Online Appendix.

Throughout the paper, we use the following notation. Let $\lambda_{\max}(M)$ and $\lambda_{\min}(M)$ denote the largest and smallest eigenvalues of the square matrix M, respectively. For a vector a, $\|a\|_2 = \sqrt{a'a}$ represents the Euclidean norm of a and for a matrix A, $\|A\|_2 = \sqrt{\lambda_{\max}(A'A)}$. Also, $\mathbb{N}$ and $\mathbb{R}^m$ signify the set of natural numbers and the set of real $m \times 1$ vectors, respectively. For two scalars a and b, $a \vee b = \max(a, b)$. Furthermore, I_p stands for the identity matrix of dimension p. Convergence in distribution is denoted by $\xrightarrow{d}$, while the abbreviation $a.s.$ stands for 'almost surely.' Finally, $a_n = o_P(1)$ denotes that the sequence a_n tends to zero in probability, $a_n = O_P(1)$ signifies that a_n

produce correct coverage when the exogeneity condition for the instruments is violated. Similar distortions arise for misspecification-robust inference (see, for example, Hall and Inoue, 2003) in the presence of weak or irrelevant instruments. A unified inference framework, that is fully robust to both model misspecification and potential lack of identification, is currently not available (Andrews, Stock and Sun, 2019, p. 749). Thus, valid pre-tests – as the test for instrument exogeneity developed in this paper – could still be quite informative about the source of misspecification or lack of identification of the model, and may lead to more efficient estimation and inference.

is bounded in probability, and $a_n << b_n$ means that a_n/b_n tends to 0 as n grows to ∞.

2 Main setup

2.1 Model, notation and drifting parameterization

Consider the linear regression model

$$y_i = \eta_0 + x_i'\theta_0 + \varepsilon_i, \quad i = 1, \dots, n. \tag{1}$$

All of the explanatory variables x_i are allowed to be endogenous and inference about the model parameters relies on instrumental variable (IV) methods with z_i denoting a vector of instruments available. We assume throughout the paper that the sample $\{(x_i, z_i, y_i) \in \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R} : i = 1, \dots, n\}$ is a triangular array of independent and identically distributed random vectors with common distribution P_n that is allowed to change with the sample size n. We also maintain that the regression error ε_i has zero mean under P_n. Our methodology requires that a subset of z_i are continuous random variables although z_i with a rich enough support will suffice. The overall instrument set can still contain discrete instruments but the basis functions, defined below, can be applied only to the subset of instruments that are continuous.

The relevance and exogeneity of the instruments play an essential role in obtaining standard inference in this linear IV setup. This paper is concerned with testing the exogeneity of the instruments irrespective of their strength. We assume that the informative part of the conditional mean of x_i given z_i can be reparameterized as local-to-zero in a way that covers the spectrum of all relevant identification features. More specifically, under P_n, we set:[4]

$$E(x_i|z_i) = \mu_x + \frac{C(z_i)}{n^\delta}, \tag{2}$$

where $\mu_x \in \mathbb{R}^p$ is the population mean of x_i, $\delta \in [0, +\infty]$ is a scalar parameter, and $C(z)$ is an arbitrary measurable function. This representation of the conditional expectation is quite flexible and varying δ in the specified ranges captures all the identification traits encountered in the literature.[5] Note also that we let the function $C(z)$ be unspecified instead of posing a linear relation as it is commonly done in the literature (e.g., Staiger and Stock, 1997).

[4]The fact that the right-hand-side of (2) depends on n is the key motivation of representing the sample as a triangular array. Expectations throughout the paper are taken under P_n although we do not make this explicit for notational convenience.

[5]As x_i lies in $\mathbb{R}^p$, we could extend this setting to account for the possibility that each component of the conditional mean is possibly local-to-zero with a specific value for δ. However, such a consideration may increase the notational burden without adding more insight. Instead, in the theory developed below, we focus on the case where δ is the same for all components and report simulation results on the more general configuration.

The setup when $\delta = 0$ corresponds to the case of strong instruments if $Var(C(z))$ is non-singular. The case $\delta \neq 0$ corresponds to semi strong/weak instruments as studied by Antoine and Renault (2009, 2012, 2020) and Dovonon, Doko Tchatoka and Aguessy (2023) for unconditional moment models. In this case, the information content of the instruments vanishes as the sample size grows, making the instruments progressively irrelevant. The case where the instruments are completely irrelevant or uninformative about the parameter θ_0 corresponds to the case $\delta = +\infty$.

We are interested in testing the null of exogeneity of the instruments :

$$H_0 : \quad E(\varepsilon_i | z_i) = 0, \quad a.s.$$

against the alternative:

$$H_1 : \quad \Pr\left(E(\varepsilon_i | z_i) = 0\right) < 1.$$

We consider the specification test for conditional moment restrictions proposed by Dovonon and Gospodinov (2024a) and investigate its properties under the null and the alternative when the explanatory variables and instruments are consistent with (1) and (2). We obtain the conditions under which this test delivers uniformly valid inference irrespective of the instruments strength; i.e., for any value of $\delta \geq 0$ while leaving $C(z)$ unspecified.

Under the null of exogeneity, ε_i is uncorrelated with any suitable measurable function of z_i. Let $\{g_l(\cdot) : \mathbb{R}^m \mapsto \mathbb{R}\}_l$ be a sequence of functions that forms a basis of the separable Hilbert space $L^2(P_z) := L^2(\mathbb{R}^m; \mathcal{B}(\mathbb{R}^m); P_z)$ of square P_z-integrable real-valued functions defined on $\mathbb{R}^m$, where P_z is the probability distribution of z_i and $\mathcal{B}(\mathbb{R}^m)$ is the Borel σ-algebra of $\mathbb{R}^m$. (We refer to de Jong and Bierens (1994), Donald, Imbens and Newey (2003), and Dovonon and Gospodinov (2024a), among others, for specific choices of g_l.)

Let $g^{(k)}(z) = (g_1(z), g_2(z), \ldots, g_k(z))'$ for $k = 1, 2, \ldots$, and $Z_i = g^{(k)}(z_i)$. We consider a test for H_0 based on the moment restriction.[6]

$$Cov(Z_i, \varepsilon_i) = E\left((Z_i - \mu_z)[y_i - \mu_y - (x_i - \mu_x)'\theta_0]\right) = 0, \quad k = 1, \ldots,$$

with a feasible version given by:

$$E\left((Z_i - \bar{Z})[y_i - \bar{y} - (x_i - \bar{x})'\theta_0]\right) = 0, \quad k = 1, \ldots, \tag{3}$$

[6] Exploiting covariance leads to a demeaned moment equality which presents the non-trivial advantage of getting rid of the intercept η. So long as the constant instrument is included, the intercept in IV regressions is always strongly identified regardless of the strength of the remaining instruments. In case of (semi)-weak instruments, the intercept is typically estimated at a faster rate than the slope parameters. Standard (non-demeaned) moment restrictions would, therefore, make the theoretical analysis more complicated because of the induced heterogeneity of rates while demeaned restrictions are immune to such issues.

where $\mu_\alpha = E(\alpha_i)$, and $\bar{\alpha} = \sum_{i=1}^n \alpha_i/n$.

Finally, letting $\hat{W}$ be a (k, k)-symmetric positive definite matrix and $\tilde{\theta}$ be the generalized method of moments (GMM) estimator of θ_0, based on this sequence of feasible moments restrictions, $\tilde{\theta}$ is given by:

$$\tilde{\theta} = \arg\min_{\theta} (\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\theta)' \hat{W} (\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\theta) = \left(\tilde{\mu}_{zx}' \hat{W} \tilde{\mu}_{zx}\right)^{-1} \left(\tilde{\mu}_{zx}' \hat{W} \tilde{\mu}_{zy}\right),$$

with $\tilde{\mu}_{\alpha\beta} = n^{-1} \sum_{i=1}^n (\alpha_i - \bar{\alpha})(\beta_i - \bar{\beta})'$, where $(\alpha_i, \beta_i) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ for $i = 1, \ldots, n$, and for some integers n_1 and n_2.

The test statistic is based on the GMM objective function using the weighting matrix $\tilde{V}^{-1}$ given by

$$\tilde{V} = n^{-1} \sum_{i=1}^{n} [y_i - \bar{y} - (x_i - \bar{x})'\tilde{\theta}]^2 (Z_i - \bar{Z})(Z_i - \bar{Z})', \tag{4}$$

where the weighting matrix $\hat{W}$ associated with $\tilde{\theta}$ is possibly non-optimal. Typically, $\hat{W}$ is set to I_k or $\hat{V}_z = n^{-1} \sum_{i=1}^n (Z_i - \bar{Z})(Z_i - \bar{Z})'$ which yields the two-stage least squares estimator (2SLS). Let $\hat{\theta}$ be the two-step GMM (2SGMM) estimator based on $\hat{W} = \tilde{V}^{-1}$,

$$\hat{\theta} = \left(\tilde{\mu}_{zx}' \tilde{V}^{-1} \tilde{\mu}_{zx}\right)^{-1} \left(\tilde{\mu}_{zx}' \tilde{V}^{-1} \tilde{\mu}_{zy}\right). \tag{5}$$

Let $\hat{V}$ be defined as $\tilde{V}$ but with $\hat{\theta}$ replacing $\tilde{\theta}$, that is:

$$\hat{V} = n^{-1} \sum_{i=1}^{n} [y_i - \bar{y} - (x_i - \bar{x})'\hat{\theta}]^2 (Z_i - \bar{Z})(Z_i - \bar{Z})', \tag{6}$$

and let

$$J_{n,k} := n \left(\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta}\right)' \hat{V}^{-1} \left(\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta}\right)$$

be *a version* of the Sargan-Hansen J test for overidentifying restrictions.[7] The test statistic for the exogeneity condition is given by:

$$S_{n,k} = \frac{J_{n,k} - k}{\sqrt{2k}}. \tag{7}$$

In developing the limiting theory, we adopt a setup with an expanding set of instruments $k \to \infty$, as a function of n.

[7]Note that the standard Sargan-Hansen test statistic, $J_{n,k}$, is defined using $\tilde{V}$ instead of $\hat{V}$. For reasons that we will explain later in Remark 1, the slightly modified version that we consider in this paper for the specification test statistic is required to obtain asymptotic uniform validity of our specification test over the whole range $\delta \in [0, +\infty]$. We index the statistic $J_{n,k}$ by k to signify that it is a function of an expanding set of k instruments, while we reserve the notation J_n for the conventional test for overidentifying restrictions with a fixed number of instruments.

2.2 Assumptions

This section collects the main assumptions that provide the basis for characterizing the limiting behavior of the GMM estimator, the weighting matrix and the test statistic for instrument exogeneity. First, the representation of the conditional mean given by (2) is formally stated in the following assumption, recalling that the expectations are under P_n.

Assumption 1 *There exists an $\mathbb{R}^p$-valued function $C(z)$ and $\delta \in [0, +\infty]$ such that, letting $a_k := E\left((Z_i - \mu_z)C(z_i)'\right)$,*

$$(a)\ E(x_i|z_i) = \mu_x + n^{-\delta}C(z_i); \quad (b)\ \|a_k\|_2 = O(1),\ \text{Rank}(a_k) = p,\ \textit{and}\ \liminf_k \lambda_{\min}(a_k'a_k) > 0.$$

Part (a) captures the possibility of varying degrees of identification. Part (b) sets the magnitude of $a_k := Cov(Z_i, C(z_i))$. Although this is a (k,p)-matrix with k growing, the order of magnitude imposed on the norm of this matrix is realistic, especially in the light of the case of linear IV models. Indeed, when $C(z_i) = \Pi'(Z_i - \mu_z)$ and $\delta = 0$ as in standard linear IV models, we have

$$x_i = \mu_x + \Pi'(Z_i - \mu_z) + v_i,$$

with $\Pi \in \mathbb{R}^{k\times p}$ and $E(v_i|z_i) = 0$. Hence,

$$a_k := E[(Z_i - \mu_z)C(z_i)'] = E[(Z_i - \mu_z)(x_i - \mu_x)'] = V_z\Pi,$$

with $V_z := E[(Z_i - \mu_z)(Z_i - \mu_z)']$. Besides,

$$V_x := E[(x_i - \mu_x)(x_i - \mu_x)'] = \Pi' V_z \Pi + Var(v_i).$$

Thus, $V_x - \lambda_{\min}(V_z)\Pi'\Pi$ is positive semidefinite. Under the condition that V_z has its smallest eigenvalue bounded away from 0, we can claim that $\|\Pi\|_2 < \infty$. Therefore,

$$\|a_k\|_2 = \left\|E\left[(Z_i - \mu_z)C(z_i)'\right]\right\|_2 = \|V_z\Pi\|_2 \leq \|V_z\|_2 \cdot \|\Pi\|_2 = O(1)$$

using that $\|V_z\|_2 = O(1)$ which is implied by Assumption 2 below. □

The requirement that $\text{Rank}(a_k) = p$, for k large enough, ensures first-order local identification of the model when $\delta \in [0, 1/2)$. This condition is not restrictive either. A sufficient condition for it to hold is that $Var\left(E(x_i|z_i)\right)$ is positive definite. For $\delta = 0$, we establish the connection as follows. By the definition of the components of Z_i as basis functions of $L^2(P_z)$, for k large enough, we write

$$E(x_i - \mu_x|z_i) \simeq \Lambda Z_i,$$

where Λ is a matrix of suitable size. Thus,

$$Var\left(E(x_i|z_i)\right) = E\left[E(x_i - \mu_x|z_i)E(x_i - \mu_x|z_i)'\right] \simeq E\left[\Lambda(Z_i - \mu_z)C(z_i)'\right] = \Lambda a_k,$$

where we use part (a) of the assumption, the fact that $E(C(z_i)) = 0$ and the definition of a_k.[8] Then, it follows that:

$$p = \text{Rank}[Var\left(E(x_i|z_i)\right)] \leq \text{Rank}(a_k) \leq p.$$

Similar to the result of Antoine and Renault (2012) for k fixed, we show that with a growing k, the model parameter is consistently estimable. The condition on the rank is useful to establish that the GMM estimator is consistent and to derive its rate of convergence. While the rank requirement implies that the singular values of a_k are non-zeros for k large enough, the last condition in part (b) rules out the possibility that a_k does have a subsequence with smallest (or more precisely, pth largest) singular values that converge to 0. This is a technical condition that is useful to evaluate the magnitude of quantities involving the inverse of $a_k'Wa_k$ in their expression.

Next, let $\tilde{Z}_i = V_z^{-1/2}(Z_i - \mu_z)$. Note that the dimensions of this matrix are allowed to grow ($k \to \infty$) which necessitates additional conditions that rule out ill-conditioned matrices.

Assumption 2 *Assume that* (a) x_i*,* ε_i*, and* Z_i *have up to finite eighth moments,* (b) *there exist* $0 < \underline{\lambda} \leq \bar{\lambda} < \infty$ *such that* $\underline{\lambda} \leq \lambda_{\min}(V_z) \leq \lambda_{\max}(V_z) \leq \bar{\lambda}$*,* (c) $\lambda_{\max}\left(E(w_i^2(Z_i - \mu_z)(Z_i - \mu_z)')\right) \leq \bar{\lambda}$*, for* $w_i \in \{\varepsilon_i, x_{hi} : h = 1, \ldots, p\}$*, and* (d) $k^{-1}\sum_{l=1}^k E(\tilde{Z}_{i,l}^8) \leq \Delta < \infty$.

This assumption proves useful in deriving and controlling the magnitude of the quadratic mean of quantities such as $\tilde{\mu}_{zx}'\hat{W}\tilde{\mu}_{zx}$ and $\tilde{\mu}_{zx}'\hat{W}\tilde{\mu}_{zy}$ that form the GMM estimator. Note also that the moment condition on Z_i in part (a) holds trivially for choices of basis functions $\{g_l(z) : l = 1, \ldots\}$ that are uniformly bounded. In this case, only the existence of fourth moments for x_i and ε_i are required. The boundedness of the eigenvalues of V_z rules out the possibilities of ill-conditioning. This is a common assumption in the literature. (See, e.g., Cattaneo, Jansson and Newey, 2018; Han and Phillips, 2006; and Dovonon and Gospodinov, 2024a.) Part (c) is not particularly restrictive and would follow from a subset of the other mild assumptions if ε and x were independent of z. Primitive conditions for the eigenvalues of V_z and quantities such as $E(w_i^2(Z_i - \mu_z)(Z_i - \mu_z)')$ to be bounded away from 0 and from above are given by Proposition S.3 of Dovonon and Gospodinov (2024b). This proposition is followed by simulations confirming, in more realistic configurations,

[8] Noting that $a_k := E[(Z_i - \mu_z)C(z_i)'] = E[(Z_i - \mu_z)(x_i - \mu_x)']$ and using the law of iterated expectations, we have $a_k \simeq V_z\Lambda'$, so that $Var\left(E(x_i|z_i)\right) \simeq \Lambda V_z \Lambda'$.

that part (c) is not too restrictive. Similar conditions on eigenvalues will subsequently be required in Assumptions 4 and 5. Parts (b) and (d) of Assumption 2 are useful to claim that, for any $1 \leq \nu \leq 4$,

$$E\left((Z_i - \mu_z)'(Z_i - \mu_z)/k\right)^{\nu} = O(1), \tag{8}$$

$\sqrt{n}(\bar{Z} - \mu_z)$ is, component-wise, asymptotically normally distributed and

$$\|\sqrt{n}(\bar{Z} - \mu_z)\|_2 = O_P(\sqrt{k}). \tag{9}$$

This last claim follows by observing that $E(\|\sqrt{n}(\bar{Z} - \mu_z)\|_2^2) = O(k)$. We rely on the properties (8) and (9) routinely in the proofs of the subsequent results.

Finally, we make the following assumption on the sequence of weighting matrices $\hat{W}$.

Assumption 3 *Assume that there exists W a nonrandom (k,k)-matrix symmetric positive definite such that $\|\hat{W} - W\|_2 = o_P(k^{-1/2})$ and $\lambda_{\max}(W) \leq \bar{\lambda}$ (with $\bar{\lambda}$ as in Assumption 2.)*

This assumption is trivially satisfied if the sequence of weighting matrix is set to I_k. Also, Proposition A.2 of Dovonon and Gospodinov (2023) establishes that the standard weighting matrix for IV estimation - given by the inverse of sample variance of the instruments - satisfies this assumption under mild conditions.

2.3 Robust choice of k

In this section, we study – under the null – the asymptotic stochastic order of magnitude of the GMM estimator $\tilde{\theta}$ for different values of $\delta \in [0, +\infty]$, and $k = o(n)$ but growing with the sample size. This analysis allows us to propose a robust choice of k that guarantees the most favorable behavior of the GMM estimator in terms of consistency and rate of convergence regardless of the degree of identification; i.e., for any value of $\delta \in [0, +\infty]$. Let $\mathbb{C}_{1k}$ be the p-vector and $\mathbb{V}_{1k}$ the (p,p)-matrix defined respectively by:

$$\mathbb{C}_{1k} := k^{-1}E\left(q_{1i} \cdot \varepsilon_i \cdot (x_i - \mu_x)\right) \quad \text{and} \quad \mathbb{V}_{1k} := k^{-1}E\left(q_{1i} \cdot (x_i - \mu_x)(x_i - \mu_x)'\right),$$

with $q_{1i} := (Z_i - \mu_z)'W(Z_i - \mu_z)$.

Theorem 2.1 *Under H_0, if Assumptions 1, 2 and 3 hold and $k \to \infty$ with $k = o(n)$, we have the following:*

(a) *If* $\delta \geq 1/2$, *then* $\tilde{\theta} = \theta_0 + \mathbb{V}_{1k}^{-1}\mathbb{C}_{1k} + O_P(k^{-1/2})$.

(b) *If* $0 \leq \delta < 1/2$, *then:*

(b1) *If* $k \ll n^{1/2-\delta}$ *or* $k \sim n^{1/2-\delta}$, $\tilde{\theta} = \theta_0 + O_P(n^{-1/2+\delta})$.

(b2) *If* $n^{1/2-\delta} \ll k \ll n^{1-2\delta}$, $\tilde{\theta} = \theta_0 + n^{2\delta-1}k(a_k' W a_k)^{-1}\mathbb{C}_{1k} + O_P\left(k^{1/2}n^{2\delta-1} \vee k^2 n^{4\delta-2}\right)$.

(b3) *If* $k \sim n^{1-2\delta}$, $\tilde{\theta} = \theta_0 + (\mathbb{V}_{1k} + a_k' W a_k)^{-1}\mathbb{C}_{1k} + O_P(n^{-1/2+\delta})$.

(b4) *If* $k \gg n^{1-2\delta}$, $\tilde{\theta} = \theta_0 + \mathbb{V}_{1k}^{-1}\mathbb{C}_{1k} + O_P\left(\frac{1}{\sqrt{k}} \vee \frac{n^{1-2\delta}}{k}\right)$.

The proof of Theorem 2.1 is provided in the Online Appendix. Part (a) of this theorem is an extension of the result of Dovonon and Gospodinov (2023) who study the case $\delta = +\infty$. They found that the GMM estimator has a probability limit but is inconsistent. Part (a) shows that this actually holds for any $\delta \geq 1/2$. Fixed number of instruments k would lead, as well known in the literature on weak instruments, to GMM estimators without probability limit (see, e.g., Staiger and Stock, 1997; Andrews and Cheng, 2012; among others).

Part (b) of this theorem is new and quite interesting. For $0 \leq \delta < 1/2$, it appears that consistent estimation is possible. However, this hinges on the choice k. If k grows at a slower rate than $n^{1-2\delta}$, then the GMM estimator is consistent. If k grows at the same rate as $n^{1-2\delta}$ or faster, then, consistency is lost but convergence to a probability limit is warranted. A further consideration of the cases of convergence - (b1) and (b2) - shows that the sharpest rate of convergence of the estimator is $n^{-1/2+\delta}$ which is obtained by the choices of $k \sim n^{1/2-\delta}$ or $k \ll n^{1/2-\delta}$.

The perverse effect of large k may be connected to the results of Newey and Smith (2004) who show that increased number of moment restrictions translates into bias for the GMM estimator. In our configuration, having k increasing too fast leads to a pervasive bias.

As the practitioner may be agnostic about the value of δ which may range from 0 to $+\infty$, a good point to address concerns how can we choose k so that consistent estimation is guaranteed regardless of the value of $\delta \in [0, 1/2[$ without altering the convergence of the estimator when $\delta \geq 1/2$.[9] The standard approach consists in choosing $k \sim n^\alpha$ for some $\alpha > 0$. Nevertheless, for a given α, it is always possible to find a range for $\delta < 1/2$ such that $k \sim n^\alpha \gg n^{1/2-\delta}$; especially for values of δ close to 1/2. Because of this, the quest for robustness points to choices of sequences k that have a slower rate of explosion than power functions. This motivates our consideration of

$$k \sim a(\log n)^b, \quad a, b > 0.$$

[9]Note that the convergence of the estimator is important for the specification test to be valid for $\delta \geq 1/2$.

Such choices of k fit with the conditions in (b1) and therefore, guarantees not only consistent estimation but also the sharpest rate of convergence when $0 \leq \delta < 1/2$ while preserving convergence when $\delta \geq 1/2$.

It is worth mentioning that the rate $n^{-1/2+\delta}$ derived for the GMM estimator corresponds to the optimal rate derived by Hahn and Kuersteiner (2002) and Antoine and Renault (2009, 2012) for the GMM when the unconditional moment restrictions are local to 0 with $0 \leq \delta < 1/2$. Theorem 2.1(b1) extends this result to increasing number of moment restrictions obtained from conditional moment restriction models.

In the subsequent development, we will consider $k \sim a(\log n)^b$, $a, b > 0$ instead of $k = o(n^{1/3})$ as in Dovonon and Gospodinov (2023) and Donald, Imbens and Newey (2003). This choice of smaller values for k may affect negatively the power of the specification test that we propose in this paper but this may be the price to pay for this test to be uniformly valid over the range of values $\delta \in [0, +\infty]$.

3 Asymptotic behavior under the null

Before deriving the asymptotic distribution of the specification test statistic, we need to shed some light on the limiting behavior of the *optimal* weighting matrix $\tilde{V}$ and the two-step GMM estimator – as given by (4) and (5), respectively – and this for all values of $\delta \in [0, +\infty]$. Let $v_i := x_i - E(x_i|z_i)$ and

$$r_{1i} = \varepsilon_i - v_i' \left(E(q_{1i} \cdot v_i v_i')\right)^{-1} \left(E(q_{1i} \cdot \varepsilon_i \cdot v_i)\right)$$

be the scaled remainder of the linear regression of $\sqrt{q_{1i}} \cdot \varepsilon_i$ on $\sqrt{q_{1i}} \cdot v_i$. (The actual remainder is $\sqrt{q_{1i}} \cdot r_{1i}$.) We use the index '1' to stress the dependence of the residual on the weighting matrix W through q_{1i}. Let

$$V_{1,\delta} = \begin{cases} E(\varepsilon_i^2 (Z_i - \mu_z)(Z_i - \mu_z)') & \text{if} \quad 0 \leq \delta < 1/2, \\ E(r_{1i}^2 (Z_i - \mu_z)(Z_i - \mu_z)') & \text{if} \quad \delta \geq 1/2. \end{cases}$$

The next result derives the probability limits of $\tilde{V}$ and $\tilde{V}^{-1}$ and that of the two-step GMM estimator. In addition to Assumptions 1, 2 and 3, we make the following assumption.

Assumption 4 *With $Z_i = g^{(k)}(z_i)$ and $\underline{\lambda}$ and $\bar{\lambda}$ defined as in Assumption 2, we have:*

(a) $k^{-2} \sum_{l,m=1}^{k} Var(U_i(Z_{il} - \mu_{zl})(Z_{im} - \mu_{zm})) \leq \Delta < \infty$, *and* $\lambda_{\max}\left(E[U_i(Z_i - \mu_z)(Z_i - \mu_z)']\right) \leq \bar{\lambda}$,

with $U_i \in \{1, |\varepsilon_i|, \varepsilon_i^2, |x_{ih}-\mu_{xh}|, |v_{ih}|, |C(z_i)_h|, (x_{ih}-\mu_{xh})\cdot(x_{ih'}-\mu_{xh'}), v_{ih}^2, C(z_i)_h^2 : h, h' = 1, \ldots, p\}$,
(b) $E(r_{1i}v_i|z_i) = 0$, (c) $\underline{\lambda} \leq \lambda_{\min}(V_{1,\delta})$.

This assumption is not overly restrictive. Part (a) ensures that no component of $U_i(Z_i-\mu_z)(Z_i-\mu_z)'$ has variance that dominates the others. This assumption is useful to establish that the sample mean of this quantity converges to its population mean at a suitable rate. (See Lemma OA.1.) The boundedness of eigenvalues in (a) and (c) is standard in the literature. Part (b) is equivalent to $E(q_{1i} \cdot r_{1i} \cdot v_i|z_i) = 0$. By definition of r_{1i}, $E(q_{1i} \cdot r_{1i} \cdot v_i) = 0$. Assumption 4(b) imposes that this zero expectation holds conditionally to z_i. Note that this assumption holds if ε_i and v_i are independent of z_i or, alternatively if $E(\varepsilon_i \cdot v_i|z_i) = E(\varepsilon_i \cdot v_i)$ and $E(v_i \cdot v_i'|z_i) = E(v_i \cdot v_i')$. The latter set of conditions are those used in Dovonon and Gospodinov (2023).

In our proofs, Assumption 4(b) is useful only for the case $\delta \geq 1/2$ and under the null hypothesis. It is also worth mentioning that for $\delta > 0$ and under the condition of the model,

$$\mathbb{C}_{1k} = k^{-1}E(q_{1i} \cdot \varepsilon_i \cdot v_i) + O(n^{-\delta}), \quad \mathbb{V}_{1k} = k^{-1}E(q_{1i} \cdot v_i \cdot v_i') + O(n^{-\delta}), \quad \text{and}$$

$$\mathbb{V}_{1k}^{-1}\mathbb{C}_{1k} = \left(E(q_{1i} \cdot v_i \cdot v_i')\right)^{-1} E(q_{1i} \cdot \varepsilon_i \cdot v_i) + O(n^{-\delta}).$$

We then have the following result.

Theorem 3.1 *Suppose Assumptions 1, 2, 3, and 4 hold and* $k \to \infty$ *with* $k \sim a(\log n)^b$, *for some* $a, b > 0$. *Then, under* H_0, *we have:*

(a) *If* $0 \leq \delta < 1/2$, $\tilde{V} - V_{1,\delta} = O_P(n^{-1/2+\delta})$ *and* $\tilde{V}^{-1} - V_{1,\delta}^{-1} = O_P(n^{-1/2+\delta})$.
If $\delta \geq 1/2$, $\tilde{V} - V_{1,\delta} = O_P(k^{-1})$ *and* $\tilde{V}^{-1} - V_{1,\delta}^{-1} = O_P(k^{-1})$.

(b) *If* $0 \leq \delta < 1/2$, $\hat{\theta} = \theta_0 + O_P(n^{-1/2+\delta})$.
If $\delta \geq 1/2$, $\hat{\theta} = \theta_0 + \mathbb{V}_{2k}^{-1}\mathbb{C}_{2k} + O_P(k^{-1/2})$,

with $\mathbb{C}_{2k}$ *and* $\mathbb{V}_{2k}$ *defined as* $\mathbb{C}_{1k}$ *and* $\mathbb{V}_{1k}$ *but with* q_{1i} *replaced by* $q_{2i} := (Z_i - \mu_z)'V_{1,\delta}^{-1}(Z_i - \mu_z)$.

Theorem 3.1 highlights some interesting features of the estimation procedure especially when it comes to the probability limit of the sample variance of the estimating function evaluated at an estimator. In relation to Theorem 2.1, the two-step GMM estimator turns out to be a particular case of the estimator $\tilde{\theta}$ associated with some weighting matrix $\hat{W}$. In particular, note that all the conditions of Theorem 2.1 are fulfilled for $\hat{\theta}$ once it is established that $\tilde{V}^{-1} - V_{1,\delta}^{-1} = o_P(k^{-1/2})$ as is done in part (a). Hence, as expected, while $\hat{\theta}$ is consistent when $0 \leq \delta < 1/2$, it is not consistent in general when $\delta \geq 1/2$. Instead, it converges to a probability limit $\theta_0 + \mathbb{V}_{2k}^{-1}\mathbb{C}_{2k}$ at the same rate $k^{-1/2}$ as in the general case in Theorem 2.1. Note however that the probability limit is different as it depends on the limit of the sequence of the weighting matrices. It is useful to mention that this dependence on the probability limit of the sequence of weighting matrices vanishes if we make the stronger assumption of Dovonon and Gospodinov (2023) which amounts to $E(\varepsilon_i v_i|z_i) = E(\varepsilon_i v_i)$ and $E(v_i v_i'|z_i) = E(v_i v_i')$.

Remark 1 *This dependence on the weighting matrix of the probability limit of $\hat{\theta}$ in the case $\delta \geq 1/2$ justifies the slight change to the definition of the $J_{n,k}$ (therefore, to the test statistic $S_{n,k}$) that we mentioned above. Theorem 3.1 shows that $\tilde{V}$ converges to $E(r_{1i}^2(Z_i - \mu_z)(Z_i - \mu_z)')$ which depends on W (the weighting matrix of the first-step GMM estimation). Besides, as we show in the proof of Theorem 3.2 below, the leading term of $\sqrt{n}(\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta})$ is $n^{-1/2}\sum_{i=1}^n r_{2i}(Z_i - \mu_z)$ (see definition of r_{2i} below) and it is important to use a sequence of weighting matrices that converges to $E(r_{2i}^2(Z_i - \mu_z)(Z_i - \mu_z)')^{-1}$ to have asymptotic normality of the test statistic $S_{n,k}$, therefore, uniform validity of the test. While we could not obtain this property using $\tilde{V}$, the asymptotic normality is obtained for $\delta \geq 1/2$ using $\hat{V}^{-1}$ as weighting matrix. This is due to the fact that the latter is obtained using residuals that are evaluated at the right estimator $\hat{\theta}$.*

We next turn to the derivation of the asymptotic distribution of the specification test statistic $S_{n,k}$. Define r_{2i} as r_{1i} but with the weighting matrix W replaced by $V_{1,\delta}^{-1}$, and $V_{2,\delta}$ as $V_{1,\delta}$ but with r_{1i} replaced by r_{2i}, that is:

$$r_{2i} = \varepsilon_i - v_i \left(E(q_{2i} \cdot v_i v_i')\right)^{-1} \left(E(q_{2i} \cdot \varepsilon_i \cdot v_i)\right), \quad \text{with} \quad q_{2i} = (Z_i - \mu_z)' V_{1,\delta}^{-1} (Z_i - \mu_z),$$

and

$$V_{2,\delta} = \begin{cases} E(\varepsilon_i^2 (Z_i - \mu_z)(Z_i - \mu_z)') & \text{if} \quad 0 \leq \delta < 1/2, \\ E(r_{2i}^2 (Z_i - \mu_z)(Z_i - \mu_z)') & \text{if} \quad \delta \geq 1/2. \end{cases}$$

Theorem 3.2 *Suppose Assumptions 1, 2, 3, and 4 hold with Assumption 4(b,c) holding for r_{2i} and $V_{2,\delta}$ as well, and $k \to \infty$ with $k \sim a(\log n)^b$, for some $a, b > 0$. Then, under H_0, for any value of $\delta \in [0, +\infty]$, we have:*

$$S_{n,k} \xrightarrow{d} N(0,1).$$

This result establishes the asymptotic uniformity of the specification testing procedure using $S_{n,k}$ over values of $\delta \in [0, +\infty]$. This main result shows that irrespective of the strength of the instruments, the proposed specification test statistic is asymptotically standard normal.

4 Asymptotic behavior under the alternative

In this section, we assume that the exogeneity of the instruments is compromised, that is:

$$H_1: \quad \Pr(E(\varepsilon_i(\eta,\theta)|z_i)=0)<1, \quad \text{for any } (\eta,\theta)\in\mathbb{R}^{p+1}, \quad \text{with } \varepsilon_i(\eta,\theta):=y_i-\eta-x_i'\theta.$$

Under H_1, Lemma 4.4 of Dovonon and Gospodinov (2024a) (see also Lemma 1 of de Jong and Bierens (1994)) ensures that, for a suitable choice of basis functions $(g^{(k)}(z))_{k\in\mathbb{N}}$ and any compact subset $\mathcal{C}$ of $\mathbb{R}^{p+1}$,

$$\exists k_0\in\mathbb{N} \text{ and } \delta_0>0: \quad \inf_{(\eta,\theta)\in\mathcal{C}}\|E(g^{(k)}(z_i)\varepsilon_i(\eta,\theta))\|_2>\delta_0.$$

With this insight, we can claim that under the alternative, for k large enough,

$$c_z:=E(g^{(k)}(z_i)\varepsilon_i(\eta_0,\theta_0)):=E(Z_i\varepsilon_i)\neq 0.$$

For the same reasons as those that led us to impose that $\|a_k\|_2:=\|E(Z_i-\mu_z)(x_i-\mu_x)'\|_2=O(1)$, we shall impose that $\|c_z\|_2=O(1)$. Indeed, if $E(\varepsilon_i|z_i)$ were a linear function of $Z_i-\mu_z$, we would have:

$$\varepsilon_i=(Z_i-\mu_z)'d+w_i, \quad \text{with} \quad E(w_i|z_i)=0, \quad \text{for } k \text{ large enough.}$$

Then, since $Var(\varepsilon_i)=d'V_zd+Var(w_i)$ and $\lambda_{\min}(V_z)$ bounded away from 0, we necessarily have $\|d\|_2=O(1)$. Hence, by the fact that $\|V_z\|_2=O(1)$, we have $c_z:=E((Z_i-\mu_z)\varepsilon_i)=V_zd=O(1)$. Thus, we maintain under the alternative that $c_z\neq 0$ and $\|c_z\|_2=O(1)$. This approach amounts to setting $k_1\sim\|c_z\|_2$ in Dovonon and Gospodinov (2023, Section 4) to be bounded instead of allowing it to grow with the sample size. Nevertheless, we reach the same conclusion as them in the case $\delta=\infty$ which is of interest in their study.

We next explore the limit behavior of $S_{n,k}$ under the alternative and show that it yields a consistent test. This will require that we first investigate the behavior under the alternative of the GMM estimator and the *optimal* weighting matrix. The limiting behavior of the first-step GMM estimator for different degrees of identification (different values of δ) is presented in Proposition OA.2 in the Online Appendix. When $\delta=0$,

$$\tilde{\theta}=\theta_0+(a_k'Wa_k)^{-1}a_k'Wc_z+O_P\left(k^{-1/2}\right)$$

so that $\tilde{\theta}$ converges to its pseudo-true value with an asymptotic bias $b_{k,0}=(a_k'Wa_k)^{-1}a_k'Wc_z$.[10]

[10] The pseudo-true value reduces to the true value θ_0 if and only if $c_z=0$; that is the model is correctly specified.

For $\delta > 0$, the GMM estimator diverges (at rate $\sqrt{n}/k$ for $\delta \geq 1/2$ and n^δ for $0 < \delta < 1/2$) to infinity which is in line with Dovonon and Gospodinov (2023) who focus on the case $\delta = \infty$.

The weighting matrix associated to $\tilde{\theta}$ is given by (4). To establish its order of magnitude we introduce the following notation and assumption. For any $u \in \mathbb{R}^p$, define

$$V_3(u) = E\left([u'(x_i - \mu_x)]^2 \cdot (Z_i - \mu_z)(Z_i - \mu_z)'\right)$$

and let

$$V_{3,0} = E\left([\varepsilon_i - b'_{k,0}(x_i - \mu_x)]^2 (Z_i - \mu_z)(Z_i - \mu_z)'\right). \tag{10}$$

For $\delta = 0$, the term $\varepsilon_i - b'_{k,0}(x_i - \mu_x)$ in $V_{3,0}$ is the leading term of the prediction error of y_i using the inconsistent estimator $\tilde{\theta}$. In what follows, $V_3(u)$ will determine the leading part of $\tilde{V}$ with u set to the estimation error $\tilde{e}$ in the cases where $\delta > 0$ while $V_{3,0}$ will be established to be the probability limit of $\tilde{V}$ in the case where $\delta = 0$.

Assumption 5 *(a)* $\inf_{\{u\in\mathbb{R}^p:\|u\|_2=1\}} \lambda_{\min}(V_3(u)) \geq \underline{\lambda} > 0$. *(b)* $\lambda_{\min}(W) \geq \underline{\lambda}$ *and* $\lambda_{\min}(V_{3,0}) \geq \underline{\lambda}$. *(c) There exists* $h \in \{1, \ldots, p\}$*:* $\lambda_{\min}\left(Var\left((x_{ih} - \mu_{xh})(Z_i - \mu_z)\right)\right) \geq \underline{\lambda}$. *(d)* $\liminf_k \|a'_k W c_z\|_2 > 0$.

Parts (a), (b) and (c) of Assumption 5 are standard and are similar to Assumptions 2(b) and 4(c). Part (d) imposes that $W^{1/2}a_k$ and $W^{1/2}c_z$ are not orthogonal for all k large enough. If we set $W = I_k$, this condition amounts to the requirement that the smallest absolute inner product of the columns of a_k by c_z is bounded away from 0 as k grows, except maybe for finitely many k. This condition ensures that the leading term of the expansion of $\tilde{\theta} - \theta_0$, as it appears in Proposition OA.2(a), does not vanish. This is useful to make a claim about the order of magnitude of $\|\tilde{\theta} - \theta_0\|_2^{-1}$.

Proposition OA.3 in the Online Appendix characterizes the properties of the optimal weighting matrix for various values of δ under H_1. For $\delta > 0$, $\tilde{V}$ diverges so that $\tilde{V}^{-1}$ shrinks to 0 as the sample size grows. But, when $\delta = 0$, $\tilde{V}$ and $\tilde{V}^{-1}$ converge to $V_{3,0}$ and $V_{3,0}^{-1}$, respectively. This peculiar behavior of $\tilde{V}$ requires that we study separately the 2SGMM estimator. In particular, Assumption 2(c) on the sequence of weighting matrices, under which Proposition OA.2 is derived, is not fulfilled by $\tilde{V}$.

For these reasons, we establish below the behavior of the 2SGMM estimator $\hat{\theta}$ and the estimated *optimal* weighting matrix $\hat{V}$ which uses the 2SGMM residuals $\hat{\varepsilon} = y_i - \bar{y} - (x_i - \bar{x})'\hat{\theta}$. (See Equation (6).) In order to do this, we need to impose extra regularity conditions through the following assumption. Define $\bar{D} := n^{-1/2} \sum_{i=1}^n (Z_i - \mu_z)v'_i$, with $v_i = x_i - E(x_i|z_i)$.

Assumption 6 *Let* $\mathcal{S}_1 := \{u \in \mathbb{R}^p : \|u\|_2 = 1\}$. *Assume that*

(a) $\liminf_k \|a_k' V_{3,0}^{-1} c_z\|_2 > 0, \quad \liminf_k \inf_{\{u \in \mathcal{S}_1\}} \|a_k' V_3(u)^{-1} c_z\|_2 > 0,$
$\liminf_k \inf_{\{u \in \mathcal{S}_1\}} \|\bar{D}' V_3(u)^{-1} c_z + a_k' V_3(u)^{-1} c_z\|_2$ *is positive with probability one, and* $\liminf_k \inf_{\{u \in \mathcal{S}_1\}} \|\bar{D}' V_3(u)^{-1} c_z\|_2$ *is positive with probability one.*

(b) $\sup_{\{u \in \mathcal{S}_1\}} \|\bar{D}' V_3(u)^{-1} c_z\|_2 = O_P(1).$

This assumption is an extension of Assumption 5. This is mainly needed to deal - under H_1 - with the expansion of $\hat{V}$ which is the *optimal* weighting matrix evaluated at the 2SGMM estimator $\hat{\theta}$. Given the leading term of $\hat{\theta} - \theta_0$ – under H_1 – for different value of δ, the need to control the magnitude of $\|\hat{\theta} - \theta_0\|_2^{-1}$ requires that some terms do not vanish, just as established by Proposition OA.3 for $\tilde{V}$. Part (a) of Assumption 6 essentially makes an explicit connection to these leading terms. Specifically, Assumption 6(a) helps to deal with the cases $\delta = 0$, $0 < \delta < 1/2$, $\delta = 1/2$ and $\delta > 1/2$, respectively. Note that all these restrictions are mild, whereas doing away with them may leads to a more complicated exposition. The purpose of Part (b) of this assumption is to control the magnitude of the leading term of the expansion of $\hat{\theta} - \theta_0$ in the case $\delta \geq 1/2$. Note that this assumption also is very mild. Indeed, for each value u in the compact set $\mathcal{S}_1$, $\bar{D}' V_3(u)^{-1} c_z = O_P(1)$. To see this, we observe that: $E\left(\bar{D}' V_3(u)^{-1} c_z\right) = 0$ and

$$E\left(c_z' V_3(u)^{-1} \bar{D} \bar{D}' V_3(u)^{-1} c_z\right) = c_z' V_3(u)^{-1} E[(Z_i - \mu_z) v_i' v_i (Z_i - \mu_z)'] V_3(u)^{-1} c_z \leq \bar{\lambda}_z' V_3(u)^{-2} c_z = O(1),$$

where we use Assumptions 4(a) and 5(a). This shows that $\|\bar{D}' V_3(u)^{-1} c_z\|_2 = O_P(1)$ for all $u \in \mathcal{S}_1$. Part (b) imposes that the supremum of this quantity over $\mathcal{S}_1$ is $O_P(1)$ which would follow trivially under asymptotic equicontinuity of the function $u \mapsto \|\bar{D}' V_3(u)^{-1} c_z\|_2$.

Let $V_{3,1}$ be defined as $V_{3,0}$ but with $b_{k,0}$ replaced by $b_{k,1} = (a_k' V_{3,0}^{-1} a_k)^{-1} a_k' V_{3,0}^{-1} c_z$, where $V_{3,0}$ is given by (10). We then have the following result.

Theorem 4.1 *Suppose Assumptions 1, 2, 3, 4(a), 5, and 6 hold, and* $k \sim a(\log n)^b$, *for some* $a, b > 0$. *Then, under* H_1, *we have:*

(a) *For* $\quad 0 < \delta < 1/2$,

$$\hat{\theta} - \theta_0 = n^{\delta} \left(a_k' V_3(\tilde{u})^{-1} a_k\right)^{-1} a_k' V_3(\tilde{u})^{-1} c_z + O_P\left(n^{2\delta - 1/2} \sqrt{k} \vee 1\right),$$

$$\hat{V} = V_3(\hat{e}) + O_P(n^{\delta}), \quad \textit{and} \quad \hat{V}^{-1} = V_3(\hat{e})^{-1} + O_P(n^{-3\delta}),$$

where $\tilde{e} = \tilde{\theta} - \theta_0, \quad \tilde{u} := \tilde{e}/\|\tilde{e}\|_2, \quad \hat{e} = \hat{\theta} - \theta_0$, *and* $\tilde{\theta}$ *is the first-step GMM estimator.*

(b) For $\delta = 0$,

$$\hat{\theta} - \theta_0 = \left(a_k' V_{3,0}^{-1} a_k\right)^{-1} a_k' V_{3,0}^{-1} c_z + O_P(1/\sqrt{k})$$

$$\hat{V} = V_{3,1} + O_P(1/\sqrt{k}), \quad \textit{and} \quad \hat{V}^{-1} = V_{3,1}^{-1} + O_P(1/\sqrt{k}).$$

(c) For $\delta \geq 1/2$,

$$\hat{\theta} - \theta_0 = \frac{\sqrt{n}}{k} \left(k^{-1/2} \bar{D}' V_3(\tilde{u})^{-1} k^{-1/2} \bar{D}\right)^{-1} \times \left(\bar{D}' V_3(\tilde{u})^{-1} c_z + n^{-\delta+1/2} a_k' V_3(\tilde{u})^{-1} c_z\right) + O_P\left(\sqrt{n}/k^{3/2}\right),$$

$$\hat{V} = V_3(\hat{e}) + O_P\left(\sqrt{n}/k\right), \quad \textit{and} \quad \hat{V}^{-1} = V_3(\hat{e})^{-1} + O_P\left(k^3/n^{3/2}\right),$$

where $\tilde{e}$, $\tilde{u}$ *and* $\hat{e}$ *are defined as in (a).*

Comparing the results in Theorem 4.1 with those in Propositions OA.2 and OA.3 (and the discussion above) reveals some similarities between the first-step and two-step GMM estimators. Both estimators have the same rate of convergence for various values of δ. While the weighting matrix converges to 0 for the 2SGMM, its scale does not seem to matter too much for the (rate) behavior of the estimator since there is a cancellation of its magnitude in the derivation process. The main difference between the first-step GMM and the 2SGMM estimators occurs for $0 < \delta < 1/2$ where the leading term of the estimation error appears to be random in the latter case and non-random for the first-step GMM.

We are now ready to explore the behavior of our test statistic under the alternative H_1. Define

$$\Sigma := V_3(\tilde{u}), \quad \text{if} \quad 0 < \delta < 1/2 \quad \text{and} \quad \Sigma := V_{3,0}, \quad \text{if} \quad \delta = 0,$$

and let

$$P_\delta = \Sigma^{-1/2} a_k (a_k' \Sigma^{-1} a_k)^{-1} a_k' \Sigma^{-1/2}, \quad \text{and} \quad \Delta_k = c_z' \Sigma^{-1/2} (I_k - P_\delta) \Sigma^{-1/2} c_z.$$

P_δ is the matrix of the orthogonal projection on the column span of $\Sigma^{-1/2} a_k$ while Δ_k is the squared-norm of the (orthogonal) projection of $\Sigma^{-1/2} c_z$ on the orthogonal of $\Sigma^{-1/2} a_k$.

As shown in Theorem 4.2 below, Δ_k is a leading term that determines the consistency of the proposed test in the case $0 \leq \delta < 1/2$ and it is essential that it does not vanish as k grows. Note that, through Σ, Δ_k depends on $u \in \mathbb{R}^p$ (with $u = b_{0,k}$ or $u = \tilde{e}/\|\tilde{e}\|$). In the Online Appendix

(Section OA.3), Lemma OA.4 demonstrates that for u being constant in k, Δ_k is a non-decreasing function of k while Lemma OA.5 exploits the fact that $b_{0,k}$ and $\tilde{e}/\|\tilde{e}\|$ lie in a (fixed) compact set, respectively, to show that Δ_k is uniformly bounded away from 0. This ensures the consistency of the test so long as there exists k_0 such that c_z does not belong to the column span of a_{k_0}.

Theorem 4.2 *Suppose Assumptions 1, 2, 3, 4(a), 5, and 6 hold, and $k \sim a(\log n)^b$, for some $a, b > 0$. Then, under H_1, we have:*

(a) For $0 < \delta < 1/2$: There exists a constant $C > 0$ such that,

$$J_{n,k} \geq n^{1-2\delta} \cdot C \cdot \Delta_k + O_P(n^{1/2-\delta}\sqrt{k} \vee n^{1-3\delta}), \quad \textit{w.p.a.1.}$$

(b) For $\delta = 0$: There exists a constant $C > 0$ such that,

$$J_{n,k} \geq n \cdot C \cdot \Delta_k + O_P(n/\sqrt{k}), \quad \textit{w.p.a.1.}$$

In both cases (a) and (b): As k grows, if c_z does not lie in the column span of a_k, then $\Delta_k > 0$ and, for a constant $C > 0$, we have:

$$S_{n,k} = \frac{J_{n,k} - k}{\sqrt{2k}} \geq \frac{n^{1-2\delta}}{\sqrt{k}} \cdot C \cdot \Delta_k + o_P\left(\frac{n^{1-2\delta}}{\sqrt{k}}\right), \quad \textit{w.p.a.1}$$

and both $J_{n,k}$ and $S_{n,k}$ diverge to $+\infty$, in probability as $n \to \infty$.

(c) For $\delta \geq 1/2$: There exists a random sequence $\pi_n \geq 0$ such that $\lim_{\epsilon \downarrow 0} \sup_n P(\pi_n \leq \epsilon) = 0$ and, with probability approaching 1,

$$J_{n,k} \geq k^2 \cdot \pi_n \cdot \|c_z\|_2^2 + O_P(k^3/\sqrt{n}), \quad S_{n,k} \geq 2^{-3/2} \cdot k^{3/2} \cdot \pi_n \cdot \|c_z\|_2^2 + O_P(k^{5/2}/\sqrt{n})$$

so that both $J_{n,k}$ and $S_{n,k}$ diverge to $+\infty$, in probability as $n \to \infty$.

Theorem 4.2 establishes that the test statistic diverges to $+\infty$ under the alternative as the sample size grows. Several remarks on these results are warranted. First, although the test statistic is standard normal under the null, it diverges only to $+\infty$ under the alternative implying that power would be maximum if a one-sided version of the test is implemented. This makes sense if we recall that the test is essentially a chi-squared test that is normalized to account for increasing degrees-of-freedom. As such, we should reject the null only when the test statistic is large positive. The testing rule shall then consist in rejecting the null if $S_{n,k}$ is larger than $q_{1-\alpha}$, the $(1-\alpha)$-quantile of the standard normal distribution, $\alpha \in (0, 1)$.

Second, for the case $\delta \geq 1/2$ where the instruments are weak or completely irrelevant, the power of the test is driven by the increasing number of generated instruments, k. If k were fixed, the statement in Part (c) of the theorem would not be sufficient to claim consistency of the test. More specifically, in the presence of weak/irrelevant instruments, 2SGMM diverges but at a rate that is moderated by k. Thus, the inverse of the optimal variance does not converge to 0 as fast as the signal part of $J_{n,k}$ diverges under H_1. This favorable trade-off is the source of power of the test. This extends the results of Dovonon and Gospodinov (2023), who focus on $\delta = +\infty$, to the range $\delta \in [1/2, +\infty]$. Further clarification of this surprising power properties of the test is given in Section OA.4 in the Online Appendix where a model with a single regressor and a completely irrelevant instrument is considered. More specifically, we study the limiting behavior of the statistic $J_{n,k} = n(\bar{\mu}_{zy} - \hat{\theta}\bar{\mu}_{zx})'\hat{V}^{-1}(\bar{\mu}_{zy} - \hat{\theta}\bar{\mu}_{zx})$ and observe, in this case, that $\hat{\theta} - \theta_0 = \hat{e} \sim \frac{\sqrt{n}}{k}h_n$, with $h_n = O_P(1)$.[11] Thus, $\hat{\theta}$ diverges but at a rate that is dampened by the growing k which is key to the consistency of our test. After substituting and rearranging terms, this implies that

$$\sqrt{n}(\bar{\mu}_{zy} - \hat{\theta}\bar{\mu}_{zx}) \sim \sqrt{n}c_z.$$

Furthermore, the leading term of $\hat{V}$ is a function of $\hat{e}^2$

$$\hat{V} \sim \hat{e}^2 E(x_i^2 Z_i Z_i'),$$

and, as a result,

$$J_{n,k} \sim k^2 \cdot h_n^{-2} \cdot c_z' E(x_i^2 Z_i Z_i')^{-1} c_z.$$

This shows that, due to the expanding k, $J_{n,k}$ grows to infinity under the alternative at a rate at least k^2 and $S_{n,k}$ consequently grows at the rate $k^{3/2}$. This explains the test's consistency in the case $\delta = +\infty$.

When $\delta \in [0, 1/2)$, the source of power is more standard as it stems from the increase of the sample size, regardless of k. The only associated restriction is that c_z shall not lie in the column span of a_k, that is the column span of $Cov(Z_i, x_i)$, infinitely often (as $k \to \infty$). An analogous restriction is standard in the GMM literature for k fixed and is often maintained to establish the evidence of power for the Sargan-Hansen specification test.

Third, although the results in Theorem 4.2 consider a fixed alternative by imposing that c_z is a non-vanishing $O(1)$ sequence, similar arguments can be used to elicit evidence of power under local alternatives. For instance, one may consider c_z to be local to 0 by stating that $h_n(z_i) := E(\varepsilon_i|z_i)$

[11] The notation '$a_n \sim b_n$' means that the leading term in the expansion of a_n is b_n.

or $c_z := E((Z_i - \mu_z)\varepsilon_i)$ tend to 0 at a certain rate. Then, $\|c_z\|_2$ may then be taken as proportional to $k^{-\alpha}$, with $0 \leq \alpha < 1/2$ if $0 \leq \delta < 1/4$ and $0 \leq \alpha < 3/4$ in the case $\delta \leq 1/2$. Even though these statements could be refined, studying power under local alternatives is beyond the scope of this paper is left for future research.

5 Simulation results

In this section, we undertake a Monte Carlo simulation experiment that assesses the empirical size and power properties of the specification test $S_{n,k}$ over a wide range of values that determine the identification strength of the instruments. We consider two simulation designs both of which have three potential instruments generated as $z_i = (z_{1i}, z_{2i}, z_{3i})' \sim \text{NID}(0, I_3)$. The first setup contains only one endogenous variable x_i. The $\{y_i, x_i\}$ sample is simulated as

$$\begin{aligned} y_i &= \theta_0 x_i + \alpha_0 z_{1i} + \varepsilon_i, \\ x_i &= \pi(\delta)' z_i + v_i, \end{aligned}$$

where $(\varepsilon_i, v_i)'$ is bivariate normal with mean zero and covariance matrix $\Omega = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}$ and $\theta_0 = 1$. The vector $\pi(\delta)$ takes the form $\pi(\delta) = (1/n^\delta, 1/n^\delta, 1/n^\delta)'$ and imposes the same loadings and identification strength on the instrument vector for various values of $\delta \geq 0$. In all experiments, the estimated model includes a constant term and, as a result, we add a vector of ones to the vector of instruments.

We present results for the $S_{n,k}$ test and the conventional test for overidentifying restrictions J_n. The moment condition in our test $S_{n,k}$ takes the form $E\left[g^{(k)}(z_i)\left(y_i - \bar{y} - \theta(x_i - \bar{x})\right)\right] = 0$ with $g^{(k)} = (g_1, \ldots, g_k)'$ with and $k = \lceil \log(n) \rceil$, where $\lceil a \rceil$ denotes the least integer greater than or equal to a. In the case $m := \text{size}(z) = 1$, the basis functions are constructed as $g_l = \cos(t_l \Psi(z_i)) + \sin(t_l \Psi(z_i))$, where $\Psi(z_i) = 2 \arctan(z_i)$.[12] For the case $m > 1$, $\Psi(\cdot)$ and $g_l(\cdot)$ are applied component-wise to z leading to $k = m \cdot \lceil \log(n) \rceil$ moment restrictions. We report results for the one-sided test $S_{n,k}$ at nominal level α, $\hat{Z} > q_{1-\alpha}$, where $q_{1-\alpha}$ denotes the $(1-\alpha)$ For the standard J_n test, we use the demeaned values of the instrument vector $(z_{1i}, z_{2i}, z_{3i})'$. With this instrument vector, the J_n test statistic is compared to the $\chi^2(2)$ critical values. The sample size is $n = 500$ and the number of Monte Carlo replications is 100,000.

[12] In our implementation of the test, we do not include the original instruments z_i along with the set of basis functions. Overall, our numerical experiments suggest that this choice (provided that z_i are transformed to be on a similar scale as $g^k(z_i)$) makes very little difference to the results.

For better visualization and more compact reporting of the results, we plot the empirical size and power curves on a grid of values that represent the identification strength of the instruments and the deviations from the null hypothesis, respectively. For the size computations (at 5%, 10%, 20%, 80%, 90% and 95% nominal levels), we set $\alpha_0 = 0$ and $\delta = (100, 5, 2, 1, 0.7, 0.5, 0.4, 0.3, 0.2, 0.1, 0)$ in $\pi(\delta)$ defined above. For the power computations (at 5% nominal level), we construct the grid $\alpha_0 = (0, 0.005, 0.01, \ldots, 0.4)$ and plot the empirical power curves for four different degrees of the identification signal: 'irrelevant' ($\delta = 100$), 'very weak' ($\delta = 1$), 'weak' ($\delta = 0.5$), 'semi-strong' ($\delta = 0.2$), and 'strong' ($\delta = 0$).

Figure 1 presents the empirical rejection rates of the $S_{n,k}$ (left plot) and J_n (right plot) tests over a range of values for the drifting sequence that span the cases of irrelevant, weak, semi-strong and strong instruments. The rejection rates of the $S_{n,k}$ are very close to the nominal levels uniformly over the different degrees of identification strength. The J_n test does not have a χ^2 limit in the part of the region that is associated with irrelevant and weak instruments. While the under-rejections of the J_n test near the origin appear small, the J_n test is inconsistent when the identification of the model is compromised, as illustrated in Figure 2.

Figure 2 plots the empirical power curves for the $S_{n,k}$ (left) and J_n (right) tests at the 5% nominal level for 5 different identification signals, parameterized by δ. Again, consistent with the evidence in Figure 1, the $S_{n,k}$ test is well-sized at the origin ($\alpha_0 = 0$) for all values of δ while the J_n test under-rejects when the identification is weak. In the weak identification cases, Figure 2 shows that the power of the J_n test plateaus at a value less than 1 and, hence, is inconsistent. By contrast, the $S_{n,k}$ test is consistent across all cases although the cases with a weaker identification signal may require larger samples (see the Online Appendix for the power curves of $S_{n,k}$ and J_n tests with $n = 5,000$). Of course, the uniformity and the robustness of the $S_{n,k}$ test comes at the cost of moderate power losses when the identification is strong. This is visible in Figures 1 and 2 where the power curve for the J_n test with strong instruments in Figure 2 is steeper and shifted to the left relative to the corresponding power curve of the $S_{n,k}$ test in Figure 1. This arises from the fact that the $S_{n,k}$ test is using more instruments than necessary, which in turn affects the power of the test.

The seemingly higher power (for moderate deviations from the null in Figure 1) of the $S_{n,k}$ test with irrelevant instruments relative to the case of strong instruments may also warrant some explanation. As discussed below Theorem 4.2, the power of the $S_{n,k}$ test when $\delta \geq 1/2$ is due to the expanding k while the source of the power in the case $\delta \in [0, 1/2)$ is more standard as it stems from the increase of the sample size n. Thus, interestingly, the power of the $S_{n,k}$ test with irrelevant instruments may exceed the power of the test with strong instruments if k is large relative to n.

The second experiment continues to use 3 potential instruments $z_i = (z_{1i}, z_{2i}, z_{3i})' \sim \text{NID}(0, I_3)$ but the model has two endogenous variables $x_i = (x_{1i}, x_{2i})'$ with the sample $\{y_i, x_i'\}$ generated as

$$\begin{aligned} y_i &= x_i'\theta_0 + z_i'\alpha_0 + \varepsilon_i, \\ x_i &= \Pi(\delta)' z_i + v_i, \end{aligned}$$

where $(\varepsilon_i, v_i')'$ is trivariate normal with mean zero and covariance matrix $\Omega = \begin{pmatrix} 1 & 0.3 & 0.3 \\ 0.3 & 1 & 0 \\ 0.3 & 0 & 1 \end{pmatrix}$ and $\theta_0 = (1,1)'$. For size and power computations, we set $\alpha_0 = (0,0,0)'$ and $\alpha_0 = (0.5, 0.5, 0.5)'$, respectively. The matrix $\Pi(\delta)$ takes the form $\Pi(\delta) = \begin{pmatrix} 1/n^{\delta_1} & 0 \\ 0 & 1/n^{\delta_3} \\ 1/n^{\delta_2} & 1/n^{\delta_4} \end{pmatrix}$ for different combinations of $(\delta_1, \delta_2, \delta_3, \delta_4)$, and is intended to illustrate the robustness of the results to heterogeneity in the drifting sequences across instruments. Again, the sample size is $n = 500$ and the number of Monte Carlo replications is 100,000.[13] The basis functions and the choice of k for the $S_{n,k}$ test are the same as in the first experiment with the data being demeaned prior to estimation.

To explore the robustness of our uniformity results to differential identification strength of the instruments, Table 1 reports the empirical size and power of the $S_{n,k}$ test for various combinations of the drifting parameters $(\delta_1, \delta_2, \delta_3, \delta_4)$ at the 1%, 5% and 10% nominal level. Panel A of Table 1 presents the results for the recommended one-sided $S_{n,k}$ test. Reassuringly, the results suggest that the proposed $S_{n,k}$ test continues to control size uniformly across different degrees of identification while it remains consistent. We should note that the derived $N(0,1)$ limit is a large-sample approximation and the sufficient condition for k that ensures uniform inference; i.e., $k = \lceil \log(500) \rceil = 7$, appears to be relatively small for the asymptotic approximation to fully kick in, especially at the more extreme tails (e.g., at 1% nominal level). These small size distortions are eliminated as n increases. They are also reduced for the two-sided test as illustrated in Panel B of Table 1. For more extensive simulations in the case of completely irrelevant instruments, see Dovonon and Gospodinov (2023).

[13] Results for $n = 100$ and non-Gaussian (multivariate t-distribution with 5 degrees of freedom) z_i and $(\varepsilon_i, v_i')'$ are reported in the Online Appendix.

6 Empirical application: Trade and economic growth

Our application revisits some results in the empirical growth literature of the impact of international trade on economic growth. Some suggestive evidence about the direction and magnitude of this impact can be obtained from a cross-country regression of income per person on some proxy of trade share (e.g., the ratio of exports or imports to GDP), controlling for other important determinants. However, proxies of trade are typically endogenous which compromises the causal interpretation of these regressions and necessitates the use of instrumental variables (for a review, see Durlauf *et al.*, 2005). For example, Frankel and Romer (1999) exploit geographical characteristics to construct a gravity-based instrument for trade openness while Hausmann *et al.* (2007) use country size (population and land area) as instruments for an indicator that captures the productivity level associated with a country's exports. But the reliability of the IV inference in identifying the causes of economic growth depends crucially on the strength and validity of the chosen instruments, as argued by Bazzi and Clemens (2013) who raised concerns about the validity of the size instruments.

We start by applying our test for instrument exogeneity to the modeling framework by Hausmann *et al.* (2007). More specifically, we consider pooled estimation of ten-year and five-year cross-country regressions for 79 countries and 3 (for ten-year regressions) or 7 (for five-year regressions) time periods for 1962–2000. The dependent variable is average annual growth in GDP per capita over the (ten-year or five-year) period. The endogenous variable of interest is the log of the initial productivity level of country's i exports, EXPY_i with control and instrumental variables including time period dummies, and logarithms of initial GDP per capita, human capital, area and population.[14] Table 2 presents results (two-step GMM estimates (with fixed and an expanding set of instruments) as well as the J_n and $S_{n,k}$ tests, along with their p-values)[15] for several combinations of log area, log population and log human capital as external instruments.

We report results for both the ten-year (Panel A) and five-year (Panel B) samples for various model specifications: column (·u) reports the GMM with fixed k while column (·c) reports the GMM results based on an expanding set of instruments k. Variables, whose cells in the table are left empty, serve as external instruments for the endogenous variable EXPY. Column (1u) replicates the results in Hausmann *et al.* (2007) although we report the results for the 2SGMM estimator while Hausmann *at al.* (2007) present the 2SLS estimates. For this specification, 'area' and 'population'

[14] The data is retrieved from the replication files of Bazzi and Clemens (2013).

[15] The choice of basis functions and tuning parameters for the $S_{n,k}$ test is the same as the one described and implemented in the simulation section. For $S_{n,k}$, we report p-values for the one-sided test.

are the excluded instruments and the J_n test strongly rejects the null of instrument exogeneity. The invalidity of the instruments jeopardizes the reliability of the statistical inference which should be adjusted using misspecification-robust standard errors as proposed by Maasoumi and Phillips (1982) and Hall and Inoue (2003). The results based on the 2SGMM estimator with an expanding k and the $S_{n,k}$ test proposed in this paper lead to similar conclusions.

The other two specifications are used here to only illustrate the advantages of our approach by including 'human capital' as an instrument in place of 'population' or 'area' in columns (2) and (3), respectively. Two main observations emerge from the results based on specifications (2) and (3) in Table 2. First, while the J_n test does not reject the null at 5% significance level in both specifications and samples, the $S_{n,k}$ test strongly suggests that the instruments are not exogenous. This may be partly due to the fact that the strength of the instruments is weaker[16] so that the asymptotic distribution of the J_n test becomes non-standard and its power using $\chi^2(1)$ critical values is compromised. Second, there are some interesting differences in the GMM estimates obtained from a fixed and an expanding set of instruments. We start by noting that when the specification tests were in agreement, as in columns (1u) and (1c), the GMM estimates exhibited very little differences across the two approaches. However, when the tests lead to different conclusions, the GMM estimates show substantial differences. The 2SGMM estimator based on an expanding set of instruments k tends to produce more stable estimates across the different specifications and across the ten-year and five-year samples. On the other hand, the 2SGMM estimates with a fixed k increase sharply relative to the first specification with a value of 0.251, for example, in specification (3u) relative to 0.092 in (1u). This behavior can be attributed to the divergence of the 2SGMM estimator under the alternative when the instruments are weak. But this rate of divergence is dampened for the 2SGMM estimator with an expanding k so that the rate for this GMM estimator is $\sqrt{n/k}$ instead of $\sqrt{n}$ as established in Section 4.

One appealing feature of our test arises in the context of just-identified models where the J_n test for overidentifying restrictions cannot be used. By contrast, our $S_{n,k}$ test can be applied to both just-identified or over-identified models, without the need of any adjustments. To demonstrate this aspect of our testing procedure, we consider the model by Frankel and Romer (1999), with the modified specifications and instruments proposed by Deij *et al.* (2019). The cross-country growth regression setup is similar to the one considered above but the main endogenous variable

[16] For specification (3·) in the last two columns of Table 2, for example, the first stage F-test has values of 2.99 and 5.70 for the ten-year and five-year samples, respectively.

is trade share measured as country i's ratio of total trade (exports + imports) to GDP. Deij *et al.* (2019) generate two alternative instruments for trade based on estimation of a bilateral trade equation on a set of geographic characteristics. The instruments are constructed by aggregating the predictions of this equation: (a) by including predictions for all potential (zero or non-zero) bilateral trade flows and (b) by including predictions only for observations with active (positive) bilateral trade. The former instrument was originally proposed by Frankel and Romer (1999) while the latter instrument is also widely used in empirical studies. For more details, see Frankel and Romer (1999) and Deij *et al.* (2019).[17]

Table 3 presents the results for the two-step GMM estimator (with expanding k) in a cross-country regression model of log of real GDP per capita on trade share (endogenous regressor) and control variables that include population, land area, distance to Equator, percentage of land in tropics, and regional dummies. The endogenous variable 'trade share' is instrumented by one of the two instruments defined above. Columns (·a) in Table 3 refer to the model that uses the instrument in part (a) while columns (·b) refer to specifications that use the instrument in part (b) that reflects only positive trade flows. These are just-identified models but our approach produces an overidentifying framework with k instruments ($k = \lceil \log(n) \rceil$ and $n = 98$).

For the most restricted specification in columns (1a) and (1b), the $S_{n,k}$ test strongly rejects the null of exogeneity of the instruments. The results for the other specifications suggest that the validity of the instruments appears to hold when more control variables are included. While the non-rejection of the null may also be attributed to the small sample size, the capability of the proposed test to assess model specification and instrument exogeneity – by converting just-identified models into models with overidentifying restrictions, irrespective of their degree of identification – proves to be quite valuable and promising.

7 Conclusion

This paper develops a framework for testing instrument exogeneity in linear IV models which is uniformly valid over the whole range of identification signal strengths. We propose a test for conditional moment restrictions with an expanding set of constructed instruments. The limiting distribution of the test is standard normal under the null and is not affected by the uncertainty about the degree of identification. We establish that the test is consistent under the alternative

[17] The data for this empirical exercise is obtained from the replication files provided by Deij *et al.* (2019).

even when the instruments are weak or completely irrelevant. This stands in contrast to the standard test for overidentifying restrictions which fails to exhibit asymptotic power when identification is compromised. Using a general drifting framework for the identification signal, we derive novel results that characterize the orders of magnitude for the GMM estimator under the null and alternative hypothesis. The proposed test is straightforward to construct and it allows the researcher to use standard inference for testing instrument exogeneity without taking a stand on whether the instruments are strong, semi-strong, weak or completely irrelevant. We illustrate the appealing properties of the test in simulations and an empirical application of the effect of trade on economic growth.

There are some interesting directions in which this work can be extended. First, it is worth exploring the properties of the version of the test statistic that uses the ordinary least squares (OLS) estimator instead of 2SGMM. The fact that OLS always has a probability limit would help control the behavior of the weighting matrix and the resulting test would be consistent. Second, the proposed framework with an expanding number of instruments can be extended to construct tests for nonlinear conditional moment restrictions that are robust to the strength of the identification signal. Finally, in order to accommodate empirical problems with limited sample sizes, it is desirable to establish the uniform validity of the bootstrap version of the proposed test for instrument exogeneity. These research topics are currently under investigation by the authors.

References

[1] Anatolyev, S. and N. Gospodinov, 2011. Specification testing in models with many instruments. Econometric Theory 27, 1–15.

[2] Andrews, D.W.K., and X. Cheng, 2012. Estimation and inference with weak, semi-strong, and strong identification. Econometrica 80, 2153–2211.

[3] Andrews, D.W.K., and J.H. Stock, 2007. Testing with many weak instruments. Journal of Econometrics 138, 24–46.

[4] Andrews, I., J.H. Stock, and L. Sun, 2019. Weak instruments in instrumental variables regression: Theory and practice. Annual Review of Economics 11, 727–753.

[5] Antoine, B., and P. Lavergne, 2023. Identification-robust nonparametric inference in a linear IV model. Journal of Econometrics 235, 1–24.

[6] Antoine, B. and E. Renault, 2009. Efficient GMM with nearly-weak instruments. The Econometrics Journal 12, S135–S171.

[7] Antoine, B. and E. Renault, 2012. Efficient minimum distance estimation with multiple rates of convergence. Journal of Econometrics 170, 350–367.

[8] Antoine, B. and E. Renault, 2020. Testing identification strength. Journal of Econometrics 218, 271–293.

[9] Bazzi, S., and M.A. Clemens, 2013. Blunt instruments: Avoiding common pitfalls in identifying the causes of economic growth. American Economic Journal: Macroeconomics 5, 152–186.

[10] Bekker, P. A., 1994. Alternative approximations to the distributions of instrumental variable estimators. Econometrica 62, 657–681.

[11] Bierens, H.J., 1982. Consistent model specification tests. Journal of Econometrics 20, 105–134.

[12] Caner, M, 2014. Near exogeneity and weak identification in generalized empirical likelihood estimators: Many moment asymptotics. Journal of Econometrics 182, 247–268.

[13] Carrasco, M., and J.P. Florens, 2000. Generalization of GMM to a continuum of moment conditions, Econometric Theory 16, 797–834.

[14] Cattaneo, M.D., M. Jansson, and W.K. Newey, 2018. Inference in linear regression models with many covariates and heteroskedasticity. Journal of the American Statistical Association 113, 1350–1361.

[15] Chao, J.C., and N.R. Swanson, 2005. Consistent estimation with a large number of weak instruments. Econometrica 73, 1673–1692.

[16] Cragg, J.G., and S.G. Donald, 1996. Testing overidentifying restrictions in unidentified models. Unpublished manuscript.

[17] Deij, S., J.B. Madsen, and L. Puzzello, 2019. When are instruments generated from geographic characteristics in bilateral relationships invalid? Journal of Applied Econometrics 36, 437–452.

[18] de Jong, R., and H.J. Bierens, 1994. On the limit behavior of a chi-square type test if the number of conditional moments tested approaches infinity. Econometric Theory 10, 70–90.

[19] Doko Tchatoka, F., and J.-M. Dufour, 2023. Exogeneity tests and weak identification in IV regressions: asymptotic theory, point estimation and identification-robust inference with exogeneity pretesting. Unpublished manuscript.

[20] Donald, S.G., G.W. Imbens and W.K. Newey, 2003. Empirical likelihood estimation and consistent tests with conditional moment restrictions. Journal of Econometrics 117, 55–93.

[21] Dovonon, P., F. Doko Tchatoka, and M. Aguessy, 2023. Relevant moment selection under mixed identification strength. Econometric Theory, forthcoming.

[22] Dovonon, P., and N. Gospodinov, 2023. Consistent specification testing with irrelevant instruments. Unpublished manuscript.

[23] Dovonon, P., and N. Gospodinov, 2024a. Specification testing for conditional moment restrictions under local identification failure. Quantitative Economics 15, 849-891.

[24] Dovonon, P., and N. Gospodinov, 2024b. Supplement to "Specification testing for conditional moment restrictions under local identification failure." Quantitative Economics Supplemental Material, 15, https://doi.org/10.3982/QE2242.

[25] Durlauf, S.N., P.A. Johnson, and J.R.W. Temple, 2005. Growth econometrics. In Handbook of Economic Growth , Vol. 1A, (eds. P. Aghion and S.N. Durlauf), Amsterdam: North-Holland, 555–677.

[26] Frandsen, B., L. Lefgren, and E. Leslie. Judging judge fixed effects. American Economic Review 113, 253–277.

[27] Frankel, J.A., and D. Romer, 1999. Does trade cause growth? American Economic Review 89, 379–399.

[28] Gospodinov, N., R. Kan, and C. Robotti, 2017. Spurious inference in reduced-rank asset-pricing models. Econometrica 85, 1613–1628.

[29] Guggenberger, P., 2012. On the asymptotic size distortion of tests when instruments locally violate the exogeneity assumption. Econometric Theory 18, 387–421.

[30] Guo, Z., H. Kang, T.T. Cai, and D.S. Small, 2018. Testing endogeneity with high dimensional covariates. Journal of Econometrics 207, 175–187.

[31] Hahn, J. and G. Kuersteiner, 2002. Discontinuities of weak instrument limiting distributions. Economics Letters 75, 325–331.

[32] Hall, A.R., and A. Inoue, 2003. The large sample behaviour of the generalized method of moments estimator in misspecified models. Journal of Econometrics 114, 361–394.

[33] Han, C., and P.C.B. Phillips, 2006. GMM with many moment conditions, Econometrica 74, 147–192.

[34] Hausmann, R., J. Hwang, and D. Rodrik, 2007. What you export matters. Journal of Economic Growth 12, 1–25.

[35] Jun, S.J., and J. Pinkse, 2009. Semiparametric tests of conditional moment restrictions under weak or partial identification. Journal of Econometrics 152, 3–18

[36] Kitamura, Y., 2006. Specification tests with instrumental variables and rank deficiency. In: Econometric Theory and Practice (D. Corbae, S.N. Durlauf, and B.E. Hansen, eds.), Cambridge University Press, 59–81.

[37] Kleibergen, F., 2005. Testing in parameters in GMM without assuming that they are identified. Econometrica 73, 1103–1123.

[38] Kolesár, M., R. Chetty, J. Friedman, E. Glaeser, and G.W. Imbens, 2015. Identification and inference with many invalid instruments. Journal of Business & Economic Statistics 33, 474–484.

[39] Maasoumi, E., and P.C.B. Phillips, 1982. On the behavior of inconsistent instrumental variable estimators. Journal of Econometrics 19, 183–201.

[40] Mikusheva, A., and L. Sun, 2022. Inference with many weak instruments. Review of Economic Studies 89, 2663–2686.

[41] Murray, M.P., 2006. Avoiding invalid instruments and coping with weak instruments. Journal of Economic Perspectives 20, 111–132.

[42] Staiger, D., and J.H. Stock, 1997. Instrumental variables regression with weak instruments. Econometrica 65, 557–586.

[43] Stock, J.H., and J.H. Wright, 2000. GMM with weak identification. Econometrica 68, 1055–1096.

[44] Tripathi, G., and Y. Kitamura, 2003. Testing conditional moment restrictions, Annals of Statistics 31, 2059–2095.

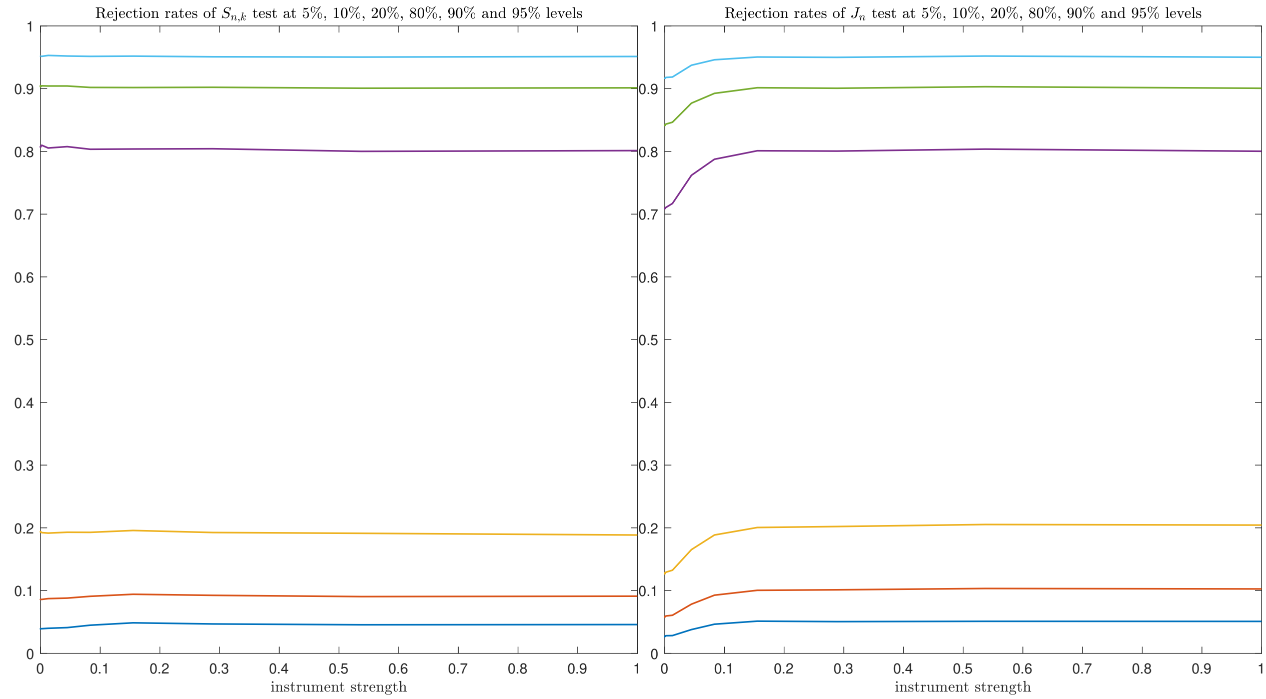

FIGURE 1. Empirical rejection rates at 5%, 10%, 20%, 80%, 90% and 95% nominal levels of the $S_{n,k}$ test (left chart) and the J_n test (right chart) as a function of the identification strength of the instruments, parameterized as $1/n^{\delta}$ for $\delta = (100, 5, 2, 1, 0.7, 0.5, 0.4, 0.3, 0.2, 0.1, 0)$.

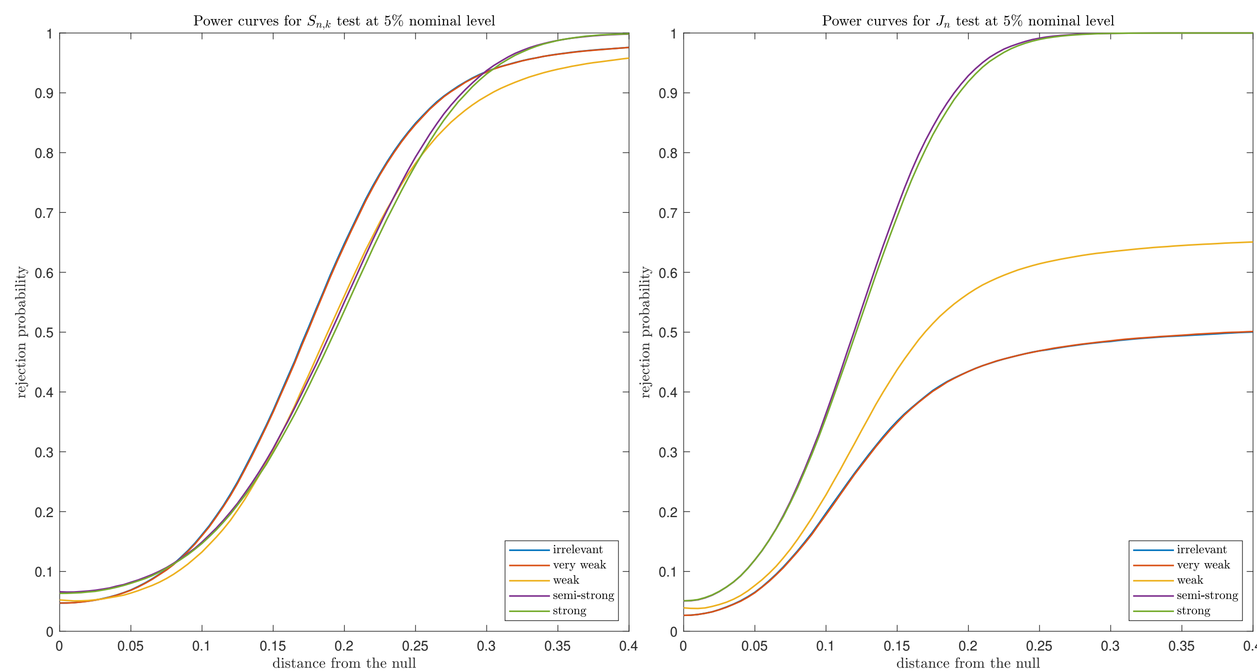

FIGURE 2. Empirical power curves at 5% nominal level of the $S_{n,k}$ test (left chart) and the J_n test (right chart) for various degrees of the identification signal: 'irrelevant' ($\delta = 100$), 'very weak' ($\delta = 1$), 'weak' ($\delta = 0.5$), 'semi-strong' ($\delta = 0.2$), and 'strong' ($\delta = 0$).

TABLE 1. Empirical rejection rates (size and power) of the $S_{n,k}$ test with instruments that exhibit differential identification strength as a function of $(\delta_1, \delta_2, \delta_3, \delta_4)$.

	Panel A: one-sided $S_{n,k}$ test						Panel B: two-sided $S_{n,k}$ test					
$(\delta_1, \delta_2, \delta_3, \delta_4)$	size			power			size			power		
	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$(0, 0.5, 0.2, 100)$	1.6	5.0	8.6	95.4	97.7	98.5	1.0	4.3	9.2	94.3	96.8	97.7
$(100, 0.3, 0.1, 100)$	1.6	4.7	8.3	89.4	93.2	94.9	1.1	4.2	9.2	87.7	91.5	93.2
$(0, 0.2, 0.5, 0)$	1.5	4.7	8.3	100	100	100	1.0	3.9	8.9	100	100	100
$(0.8, 0.2, 0.5, 0.4)$	1.4	4.2	7.4	93.8	96.1	97.1	0.9	4.3	9.7	92.8	95.1	96.1
$(0.5, 0.4, 0.3, 0.1)$	1.2	3.9	6.8	96.1	97.5	98.2	0.8	4.0	9.3	95.4	96.9	97.5
$(0, 100, 100, 0)$	1.5	4.8	8.3	100	100	100	1.0	4.0	8.9	100	100	100
$(0.1, 0.2, 0.5, 0.5)$	1.2	3.9	6.8	86.9	91.2	93.1	0.8	3.8	9.2	85.1	89.3	91.3
$(0.6, 0.5, 0.2, 1)$	1.2	3.9	7.0	89.8	93.2	94.7	0.8	4.0	9.4	88.4	91.7	93.2

Table 2. Two-step GMM estimates and specification tests for the ten-year and five-year cross-country regressions for the period 1962–2000 (Hausmann *et al.*, 2007).

	(1u)	(1c)	(2u)	(2c)	(3u)	(3c)
Panel A: ten-year sample ($n = 299$)						
log EXPY	0.092	0.092	0.132	0.074	0.251	0.080
log initial GDP/capita	-0.038	-0.038	-0.054	-0.028	-0.105	-0.031
log human capital	0.004	0.004				
log area			-0.003	-0.002		
log population					-0.009	-0.000
J_n test	11.25		0.369		0.453	
(p-value)	(0.001)		(0.544)		(0.501)	
$S_{n,k}$ test		2.628		3.491		3.255
(p-value)		(0.004)		(0.000)		(0.001)
Panel B: five-year sample ($n = 604$)						
	(1u)	(1c)	(2u)	(2c)	(3u)	(3c)
log EXPY	0.074	0.052	0.116	0.075	0.213	0.087
log initial GDP/capita	-0.030	-0.021	-0.047	-0.029	-0.088	-0.034
log human capital	0.003	0.004				
log area			-0.003	-0.003		
log population					-0.008	-0.001
J_n test	15.79		0.163		2.855	
(p-value)	(0.000)		(0.686)		(0.091)	
$S_{n,k}$ test		2.473		2.786		3.438
(p-value)		(0.007)		(0.003)		(0.000)

Table 3. Two-step GMM estimates (with expanding k) and specification test for instrument exogeneity for various specifications of the effect of trade on growth (Frankel and Romer, 1999; Deij *et al.*, 2019).

	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
trade share	1.131	1.879	0.653	0.909	0.690	0.840	0.810	0.962
log population	0.295	0.395	0.105	0.130	0.092	0.109	-0.027	-0.010
log area	-0.117	-0.028	-0.103	-0.031	-0.114	-0.069	0.070	0.085
distance to Equator			3.995	4.084				
% land in tropics					-1.563	-1.630		
Sub-Saharan Africa							-1.993	-2.010
East Asia							-0.637	-0.596
Latin America							-0.590	-0.744
$S_{n,k}$ test	5.009	3.258	1.280	0.456	1.655	0.766	0.608	0.107
(p-value)	(0.000)	(0.001)	(0.100)	(0.324)	(0.049)	(0.222)	(0.272)	(0.457)

A Appendix: Proofs of main results

Proof of Theorem 3.1: (a) We have

$$\tilde{V} = \frac{1}{n}\sum_{i=1}^{n} \tilde{\varepsilon}_i^2 (Z_i - \bar{Z})(Z_i - \bar{Z})', \quad \tilde{\varepsilon}_i = y_i - \bar{y} - (x_i - \bar{x})'\tilde{\theta}.$$

Let $b_k := 0$ if $0 \leq \delta < 1/2$ and $b_k := \mathbb{V}_{1k}^{-1}\mathbb{C}_{1k}$ if $\delta \geq 1/2$. Let $\tilde{e} = \tilde{\theta} - \theta_0 - b_k$. From Theorem 2.1,

$$\tilde{e} = O_P(1/\sqrt{k}) \text{ if } \delta \geq 1/2 \quad \text{and} \quad \tilde{e} = O_P(n^{-1/2+\delta}) \text{ if } 0 \leq \delta < 1/2.$$

Let $u_i := \varepsilon_i - (x_i - \mu_x)'b_k$. Straightforward calculations yield:

$$\tilde{\varepsilon}_i = u_i - (x_i - \mu_x)'\tilde{e} + (\bar{x} - \mu_x)'\tilde{e} + (\bar{x} - \mu_x)'b_k - \bar{\varepsilon}.$$

We have:

$$\tilde{V} = \frac{1}{n}\sum_{i=1}^{n} \tilde{\varepsilon}_i^2 (Z_i - \mu_z)(Z_i - \mu_z)' - \frac{1}{n}\sum_{i=1}^{n} \tilde{\varepsilon}_i^2 (Z_i - \mu_z)(\bar{Z} - \mu_z)' - (\bar{Z} - \mu_z)\cdot \frac{1}{n}\sum_{i=1}^{n} \tilde{\varepsilon}_i^2 (Z_i - \mu_z)'$$
$$+ \frac{1}{n}\sum_{i=1}^{n} \tilde{\varepsilon}_i^2 (\bar{Z} - \mu_z)(\bar{Z} - \mu_z)' = \frac{1}{n}\sum_{i=1}^{n} \tilde{\varepsilon}_i^2 (Z_i - \mu_z)(Z_i - \mu_z)' + O_P\left(\frac{k}{\sqrt{n}}\right).$$

(To obtain the order of magnitude, we use the fact that

$$\tilde{\varepsilon}_i^2 \leq 6\left(\varepsilon_i^2 + \|x_i - \mu_x\|_2^2 \cdot \|b_k\|_2^2 + \|x_i - \mu_x\|_2^2 \cdot \|\tilde{e}\|_2^2 + \|\bar{x} - \mu_x\|_2^2 \cdot \|\tilde{e}\|_2^2 + \|\bar{x} - \mu_x\|_2^2 \cdot \|b_k\|_2^2 + \bar{\varepsilon}^2\right).$$

Then, we use the law of large numbers and the orders of $\bar{Z} - \mu_z$, $\bar{x} - \mu_x$ and $\bar{\varepsilon}$ - see Equation (9) - to conclude.) Also,

$$\tilde{\varepsilon}_i^2 = u_i^2 + ((x_i - \mu_x)'\tilde{e})^2 + ((\bar{x} - \mu_x)'\tilde{e})^2 + ((\bar{x} - \mu_x)'b_k)^2 + \bar{\varepsilon}^2 - 2u_i(x_i - \mu_x)'\tilde{e} + 2u_i(\bar{x} - \mu_x)'\tilde{e}$$
$$+ 2u_i(\bar{x} - \mu_x)'b_k - 2u_i\bar{\varepsilon} - 2(x_i - \mu_x)'\tilde{e}\cdot(\bar{x} - \mu_x)'\tilde{e} - 2(x_i - \mu_x)'\tilde{e}\cdot(\bar{x} - \mu_x)'b_k$$
$$+ 2(x_i - \mu_x)'\tilde{e}\cdot\bar{\varepsilon} + 2(\bar{x} - \mu_x)'\tilde{e}\cdot(\bar{x} - \mu_x)'b_k - 2(\bar{x} - \mu_x)'\tilde{e}\cdot\bar{\varepsilon} - 2(\bar{x} - \mu_x)'b_k\cdot\bar{\varepsilon}$$

and, letting $\dot{Z}_i := Z_i - \mu_z$, we have:

1.

$$\frac{1}{n}\sum_{i=1}^{n}((x_i - \mu_x)'\tilde{e})^2 \dot{Z}_i\dot{Z}_i' \leq \|\tilde{e}\|_2^2 \cdot \frac{1}{n}\sum_{i=1}^{n}\|x_i - \mu_x\|_2^2 \cdot \dot{Z}_i\dot{Z}_i'$$
$$= \|\tilde{e}\|_2^2 \cdot \sum_{h=1}^{p}\left(\frac{1}{n}\sum_{i=1}^{n}(x_{ih} - \mu_{xh})^2 \cdot \dot{Z}_i\dot{Z}_i'\right) = O_P\left(\|\tilde{e}\|_2^2\right),$$

where, thanks to Assumption 4, we use Lemma OA.1 to claim that

$$n^{-1}\sum_{i=1}^{n}(x_{ih} - \mu_{xh})^2 \cdot \dot{Z}_i\dot{Z}_i' = E\left((x_{ih} - \mu_{xh})^2 \cdot \dot{Z}_i\dot{Z}_i'\right) + O_P(k/\sqrt{n}) = O_P(1).$$

(We use this Lemma and Assumption for such justification routinely through the end of the proof.)

2.
$$\frac{1}{n}\sum_{i=1}^{n}((\bar{x}-\mu_x)'\tilde{e})^2\cdot\dot{Z}_i\dot{Z}_i'\le\|\bar{x}-\mu_x\|_2^2\cdot\|\tilde{e}\|_2^2\cdot\frac{1}{n}\sum_{i=1}^{n}\dot{Z}_i\dot{Z}_i'=O_P\left(n^{-1}\|\tilde{e}\|_2^2\right).$$

3.
$$\frac{1}{n}\sum_{i=1}^{n}((\bar{x}-\mu_x)'b_k)^2\cdot\dot{Z}_i\dot{Z}_i'\le\|\bar{x}-\mu_x\|_2^2\cdot\|b_k\|_2^2\cdot\frac{1}{n}\sum_{i=1}^{n}\dot{Z}_i\dot{Z}_i'=O_P\left(\frac{1}{n}\right).$$

4.
$$\frac{1}{n}\sum_{i=1}^{n}\bar{\varepsilon}^2\cdot\dot{Z}_i\dot{Z}_i'=\bar{\varepsilon}^2\cdot\frac{1}{n}\sum_{i=1}^{n}\dot{Z}_i\dot{Z}_i'=O_P\left(\frac{1}{n}\right).$$

5. Recalling that $u_i=\varepsilon_i-(x_i-\mu_x)'b_k$, we have:

$$\left\|\frac{1}{n}\sum_{i=1}^{n}u_i(\bar{x}-\mu_x)'\tilde{e}\cdot\dot{Z}_i\dot{Z}_i'\right\|_2$$
$$\le\|\bar{x}-\mu_x\|_2\cdot\|\tilde{e}\|_2\cdot\left[\left\|\frac{1}{n}\sum_{i=1}^{n}|\varepsilon_i|\dot{Z}_i\dot{Z}_i'\right\|_2+\|b_k\|_2\cdot\sum_{h=1}^{p}\left\|\frac{1}{n}\sum_{i=1}^{n}|x_{ih}-\mu_{xh}|\dot{Z}_i\dot{Z}_i'\right\|_2\right]=O_P\left(\frac{1}{\sqrt{n}}\|\tilde{e}\|_2\right).$$

Similar derivations yield

$$\left\|\frac{1}{n}\sum_{i=1}^{n}u_i\cdot(\bar{x}-\mu_x)'b_k\cdot\dot{Z}_i\dot{Z}_i'\right\|_2=O_P\left(\frac{1}{\sqrt{n}}\right)\text{ and }\left\|\frac{1}{n}\sum_{i=1}^{n}u_i\cdot\bar{\varepsilon}\cdot\dot{Z}_i\dot{Z}_i'\right\|_2=O_P\left(\frac{1}{\sqrt{n}}\right).$$

6.

$$\left\|\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_x)'\tilde{e}\cdot(\bar{x}-\mu_x)'\tilde{e}\cdot\dot{Z}_i\dot{Z}_i'\right\|_2$$
$$\le\|\tilde{e}\|_2^2\cdot\|\bar{x}-\mu_x\|_2\cdot\left[\sum_{h=1}^{p}\left\|\frac{1}{n}\sum_{i=1}^{n}|x_{ih}-\mu_{xh}|\dot{Z}_i\dot{Z}_i'\right\|_2\right]=O_P\left(n^{-1/2}\|\tilde{e}\|_2^2\right).$$

Similarly, we obtain:

$$\left\|\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_x)'\tilde{e}\cdot(\bar{x}-\mu_x)'b_k\cdot\dot{Z}_i\dot{Z}_i'\right\|_2=O_P\left(\frac{\|\tilde{e}\|_2}{\sqrt{n}}\right),$$

and

$$\left\|\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_x)'\tilde{e}\cdot\bar{\varepsilon}\cdot\dot{Z}_i\dot{Z}_i'\right\|_2=O_P\left(\frac{\|\tilde{e}\|_2}{\sqrt{n}}\right).$$

7.

$$\left\|\frac{1}{n}\sum_{i=1}^{n}(\bar{x}-\mu_x)'\tilde{e}\cdot(\bar{x}-\mu_x)'b_k\cdot\dot{Z}_i\dot{Z}_i'\right\|_2\le\|\bar{x}-\mu_x\|_2^2\cdot\|\tilde{e}\|_2\cdot\|b_k\|_2\cdot\left\|\frac{1}{n}\sum_{i=1}^{h}\dot{Z}_i\dot{Z}_i'\right\|_2=O_P\left(\frac{\|\tilde{e}\|_2}{n}\right).$$

Similarly,

$$\left\|\frac{1}{n}\sum_{i=1}^{n}(\bar{x}-\mu_x)'\tilde{e}\cdot\bar{\varepsilon}\cdot\dot{Z}_i\dot{Z}_i'\right\|_2=O_P\left(\frac{\|\tilde{e}\|_2}{n}\right),\text{ and }\left\|\frac{1}{n}\sum_{i=1}^{n}(\bar{x}-\mu_x)'b_k\cdot\bar{\varepsilon}\cdot\dot{Z}_i\dot{Z}_i'\right\|_2=O_P\left(\frac{1}{n}\right).$$

I. Consider the case $0\leq\delta<1/2$. Using the orders derived above, we claim that

$$\begin{aligned}\tilde{V}&=\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2\dot{Z}_i\dot{Z}_i'-\frac{2}{n}\sum_{i=1}^{n}\varepsilon_i(x_i-\mu_x)'\tilde{e}\dot{Z}_i\dot{Z}_i'+O_P\left(n^{-1+2\delta}\vee n^{-1/2}\right)\\&=E\left(\varepsilon_i^2\dot{Z}_i\dot{Z}_i'\right)-\frac{2}{n}\sum_{i=1}^{n}\varepsilon_i(x_i-\mu_x)'\tilde{e}\dot{Z}_i\dot{Z}_i'+O_P\left(\frac{k}{\sqrt{n}}\right)+O_P\left(n^{-1+2\delta}\vee n^{-1/2}\right).\end{aligned}$$

We have:

$$\begin{aligned}\left\|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i(x_i-\mu_x)'\tilde{e}\dot{Z}_i\dot{Z}_i'\right|\|_2&\leq\|\tilde{e}\|_2\cdot\left\|\frac{1}{n}\sum_{i=1}^{n}|\varepsilon_i|\cdot\|x_i-\mu_x\|_2\dot{Z}_i\dot{Z}_i'\right\|_2\\&\leq\|\tilde{e}\|_2\cdot\left(\left\|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2\cdot\dot{Z}_i\dot{Z}_i'\right\|_2+\left\|\frac{1}{n}\sum_{i=1}^{n}\|x_i-\mu_x\|_2^2\cdot\dot{Z}_i\dot{Z}_i'\right\|_2\right)=O_P\left(n^{-1/2+\delta}\right).\end{aligned}$$

Thus, $\tilde{V}=V_{1,\delta}+O_P(k/\sqrt{n})+O_P(n^{-1/2+\delta})$. Since k grows slower than any (positive) power of n, we can claim that

$$\hat{V}=V_{1,\delta}+O_P(n^{-1/2+\delta}).\tag{A.1}$$

II. Consider the case $\delta\geq1/2$.

$$\begin{aligned}\hat{V}&=\frac{1}{n}\sum_{i=1}^{n}u_i^2\dot{Z}_i\dot{Z}_i'-\frac{2}{n}\sum_{i=1}^{n}u_i(x_i-\mu_x)'\tilde{e}\dot{Z}_i\dot{Z}_i'+O_P\left(\frac{1}{k}\right)\\&=E\left(u_i^2\dot{Z}_i\dot{Z}_i'\right)-\sum_{h=1}^{n}\tilde{e}_h\cdot\frac{2}{n}\sum_{i=1}^{n}u_i(x_{ih}-\mu_{xh})\dot{Z}_i\dot{Z}_i'+O_P\left(\frac{k}{\sqrt{n}}\right)+O_P\left(\frac{1}{k}\right).\end{aligned}$$

Recall that $u_i=\varepsilon_i-(x_i-\mu_x)'\mathbb{V}_{1k}^{-1}\mathbb{C}_{1k}$. Using the fact that $x_i-\mu_x=n^{-\delta}C(z_i)+v_i$, we obtain:

$$\mathbb{C}_{1k}=k^{-1}E(q_{1i}\cdot\varepsilon_i\cdot v_i)+O(n^{-\delta}),\quad\text{and}\quad\mathbb{V}_{1k}=k^{-1}E(q_{1i}\cdot v_iv_i')+O(n^{-\delta})$$

and, as a result,

$$b_k:=\mathbb{V}_{1k}^{-1}\mathbb{C}_{1k}=\left(k^{-1}E(q_{1i}\cdot v_iv_i')\right)^{-1}\left(k^{-1}E(q_{1i}\cdot\varepsilon_i\cdot v_i)\right)+O(n^{-\delta}):=b_{0k}+O(n^{-\delta}).$$

Also,

$$u_i=\varepsilon_i-(n^{-\delta}C(z_i)+v_i)'(b_{0k}+O(n^{-\delta}))=r_{1i}-n^{-\delta}C(z_i)'b_{0k}-n^{-\delta}C(z_i)'O(n^{-\delta})-v_i'O(n^{-\delta}),$$

with $r_{1i} = \varepsilon_i - v_i' b_{0k}$. Thus, we obtain that:

$$\frac{1}{n}\sum_{i=1}^{n} u_i^2 \dot{Z}_i \dot{Z}_i' = \frac{1}{n}\sum_{i=1}^{n} r_{1i}^2 \dot{Z}_i \dot{Z}_i' + O_P(n^{-\delta}) = E(r_{1i}^2 \dot{Z}_i \dot{Z}_i') + O_P(k/\sqrt{n}) + O_P(n^{-\delta}).$$

Besides,

$$(x_{ih} - \mu_{xh})u_i = (n^{-\delta} C(z_i)_h + v_{ih})[r_{1i} - n^{-\delta} C(z_i)' b_{0k} - n^{-\delta} C(z_i)' O(n^{-\delta}) - v_i' O(n^{-\delta})]$$

and it is not hard to find that:

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(x_{ih} - \mu_{xh})u_i \cdot \dot{Z}_i \dot{Z}_i' &= \frac{1}{n}\sum_{i=1}^{n} r_{1i} v_{ih} \cdot \dot{Z}_i \dot{Z}_i' + O_P(n^{-\delta}) \\ &= E(r_{1i} v_{ih} \cdot \dot{Z}_i \dot{Z}_i') + O_P(k/\sqrt{n}) + O_P(n^{-\delta}) = O_P(k/\sqrt{n}),\end{aligned}$$

where we use the fact, from Assumption 4(b), that $E(r_{1i} v_i | z_i) = 0$. It follows that

$$\tilde{V} = \frac{1}{n}\sum_{i=1}^{n} \tilde{\varepsilon}_i^2 (Z_i - \bar{Z})(Z_i - \bar{Z})' = E(r_{1i}^2 \dot{Z}_i \dot{Z}_i') + O_P(k/\sqrt{n}) + O_P(k^{-1}) := V_{1,\delta} + O_P(k^{-1}). \quad \text{(A.2)}$$

This completes the proof of part (a).

(b) This follows readily from (A.1) and (A.2) by observing that $V_{1,\delta}$ is nonsingular.

(c) Part (b) puts us in the conditions of Theorem 2.1 and the fact that $k \sim a(\log n)^b$ ensures that we are in the context of (b1) when $0 \leq \delta < 1/2$. The definition of the probability limit when $\delta \geq 1/2$ - case (a) of the theorem - is obtained with the limit of the estimated weighting matrix, $V_{1,\delta}^{-1}$. The result follows. □

Proof of Theorem 3.2: Note that, under the conditions of the theorem, using Theorem 3.1(b), we have: if $0 \leq \delta < 1/2$,

$$\|\hat{V}^{-1} - V_{2,\delta}^{-1}\|_2 = O_P(n^{-1/2+\delta}) = o_P(k^{-1})$$

and if $\delta \geq 1/2$,

$$\|\hat{V}^{-1} - V_{2,\delta}^{-1}\|_2 = O_P(k^{-1}).$$

Recall that

$$\begin{aligned}J_{n,k} &= n \cdot \left(\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta}\right)' \hat{V}^{-1} \left(\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta}\right) \\ &= n \cdot \left(\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta}\right)' V_{2,\delta}^{-1} \left(\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta}\right) + n \cdot \left(\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta}\right)' \left(\hat{V}^{-1} - V_{2,\delta}^{-1}\right) \left(\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta}\right) \\ &:= (a) + (b).\end{aligned}$$

We first show that $(b) = O_P(1)$ and therefore negligible. We have:

$$\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta} = \bar{\mu}_{z\varepsilon} - \bar{\mu}_{zx}(\hat{\theta} - \theta_0) + O_P(\sqrt{k}/n) = O_P(\sqrt{k/n}) + O_P(n^{-\delta} \vee \sqrt{k/n})O_P(\|\hat{\theta} - \theta_0\|_2).$$

From Theorem 3.1(b), we have $\hat{\theta} - \theta_0 = O_P(n^{-1/2+\delta})$ if $0 \leq \delta < 1/2$ and $\hat{\theta} - \theta_0 = O_P(1)$ if $\delta \geq 1/2$. Thus, in both cases,

$$\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta} = O_P(\sqrt{k/n}).$$

It follows that:

$$n \cdot \left| \left(\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta}\right)' \left(\hat{V}^{-1} - V_{2,\delta}^{-1}\right) \left(\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta}\right) \right| \leq n \cdot \|\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta}\|_2^2 \cdot \|\hat{V}^{-1} - V_{2,\delta}^{-1}\|_2 = O_P(1).$$

Thus,

$$J_{n,k} = n \cdot \left(\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta}\right)' V_{2,\delta}^{-1} \left(\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta}\right) + O_P(1).$$

Using the expression of $\hat{\theta} - \theta_0$, we write:

$$\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta} = \bar{\mu}_{z\varepsilon} - \bar{\mu}_{zx}(\tilde{\mu}_{zx}'\tilde{V}^{-1}\tilde{\mu}_{zx})^{-1}(\tilde{\mu}_{zx}'\tilde{V}^{-1}\tilde{\mu}_{z\varepsilon}) + O_P(\sqrt{k}/n).$$

I. Consider the case $0 \leq \delta < 1/2$. We write:

$$\begin{aligned} J_{n,k} = n \Big(&\bar{\mu}_{z\varepsilon}' V_{2,\delta}^{-1} \bar{\mu}_{z\varepsilon} + (\tilde{\mu}_{z\varepsilon}'\tilde{V}^{-1}\tilde{\mu}_{zx})(\tilde{\mu}_{zx}'\tilde{V}^{-1}\tilde{\mu}_{zx})^{-1}(\bar{\mu}_{zx}' V_{2,\delta}^{-1}\bar{\mu}_{zx})(\tilde{\mu}_{zx}'\tilde{V}^{-1}\tilde{\mu}_{zx})^{-1}(\tilde{\mu}_{zx}'\tilde{V}^{-1}\tilde{\mu}_{z\varepsilon}) \\ &- 2\,(\bar{\mu}_{z\varepsilon}' V_{2,\delta}^{-1}\bar{\mu}_{zx})(\tilde{\mu}_{zx}'\tilde{V}^{-1}\tilde{\mu}_{zx})^{-1}(\tilde{\mu}_{zx}'\tilde{V}^{-1}\tilde{\mu}_{z\varepsilon})\Big) + O_P(1) := n\bar{\mu}_{z\varepsilon}' V_{2,\delta}^{-1}\bar{\mu}_{z\varepsilon} + (1) + (2) + O_P(1). \end{aligned}$$

Using (OA.6), (OA.12), and (OA.14) from the Online Appendix, we can see that $(1) = O_P(1)$ and $(2) = O_P(1)$. Hence,

$$J_{n,k} = n\bar{\mu}_{z\varepsilon}' V_{2,\delta}^{-1}\bar{\mu}_{z\varepsilon} + O_P(1).$$

We apply Lemma B2 of Dovonon and Gospodinov (2023) with $r_i = \varepsilon_i$ and claim that

$$\frac{n\bar{\mu}_{z\varepsilon}' V_{2,\delta}^{-1}\bar{\mu}_{z\varepsilon} - k}{\sqrt{2k}} \xrightarrow{d} N(0,1).$$

This shows that $S_{n,k} \xrightarrow{d} N(0,1)$.

II. Consider the case $\delta \geq 1/2$. From Theorem 3.1(b), we have

$$\hat{\theta} - \theta_0 = \mathbb{V}_{2k}^{-1}\mathbb{C}_{2k} + \tilde{e} = (E(q_{2i}v_i v_i'))^{-1}E(q_{2i}v_i\varepsilon_i) + O(n^{-\delta}) + \tilde{e} := b_{0k} + O(n^{-\delta}) + \tilde{e},$$

with $\tilde{e} = O_P(k^{-1/2})$. Hence,

$$\begin{aligned} \tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta} &= \bar{\mu}_{z\varepsilon} - \bar{\mu}_{zx}(\hat{\theta} - \theta_0) + O_P(\sqrt{k}/n) = \bar{\mu}_{z\varepsilon} - \bar{\mu}_{zx}(b_{0k} + O(n^{-\delta}) + \tilde{e}) + O_P(\sqrt{k}/n) \\ &= \bar{\mu}_{zr_2} - \bar{\mu}_{zx}\tilde{e} + O_P(n^{-\delta}\sqrt{k/n}) + O_P(\sqrt{k}/n) = \bar{\mu}_{zr_2} - \bar{\mu}_{zx}\tilde{e} + O_P(\sqrt{k}/n), \end{aligned}$$

with $r_{2i} = \varepsilon_i - v_i' b_{0k}$. Thus,

$$\sqrt{n}(\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta}) = \sqrt{n}\bar{\mu}_{zr_2} - \sqrt{n}\bar{\mu}_{zx}\tilde{e} + O_P(\sqrt{k/n}) := A_n + B_n + O_P(\sqrt{k/n})$$

and

$$J_{n,k} = A_n' V_{2,\delta}^{-1} A_n + B_n' V_{2,\delta}^{-1} B_n + 2A_n' V_{2,\delta}^{-1} B_n + O_P(k/\sqrt{n}). \tag{A.3}$$

We have:

(1)

$$\|B_n\|_2 \le \|\sqrt{n}\bar{\mu}_{zx}\|_2 \|\tilde{e}\|_2 = O_P(n^{1/2-\delta} \vee \sqrt{k}) O_P(k^{-1/2}) = O_P(1).$$

Thus,

$$B_n' V_{2,\delta}^{-1} B_n = O_P(1). \tag{A.4}$$

(2) Again, the conditions of Lemma B2 of Dovonon and Gospodinov (2023) apply here and we have:

$$\frac{A_n' V_{2,\delta}^{-1} A_n - k}{\sqrt{2k}} \xrightarrow{d} N(0,1). \tag{A.5}$$

(3) Let us now consider $A_n' V_{2,\delta}^{-1} B_n$. We have:

$$A_n' V_{2,\delta}^{-1} B_n = \frac{1}{n} \sum_{i,j=1}^{n} r_{2i}(Z_i - \mu_z)' V_{2,\delta}^{-1} (Z_j - \mu_z)(x_j - \mu_x)' \tilde{e} := C_n \tilde{e} \tag{A.6}$$

Under the null hypothesis and Assumption 4(b), we can claim that $E(r_{2i}|z_i) = 0$ and $E(r_{2i}x_i|z_i) = 0$ so that $E(C_n) = 0$.

Similar to the proof of Theorem 3.3 of Dovonon and Gospodinov (2023), letting C_{nh} be the h-th component of C_h, we have:

$$E(C_{n,h}^2) = \frac{1}{n^2} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \sum_{i_3=1}^{n} \sum_{i_4=1}^{n} E(r_{2i_1} Z_{i_1}' V_{2,\delta}^{-1} Z_{i_2} x_{i_2 h} \cdot r_{2i_4} Z_{i_3}' V_{2,\delta}^{-1} Z_{i_4} x_{i_4 h}),$$

where we assume, without loss of generality, that $\mu_z = 0$ and $\mu_x = 0$. We have the following possibilities:

The four indices are pairwise different: Contribution to expectation is 0.
Two of the indices are equal and different from the other two:
(i_1, i_1, i_3, i_4) ($i_1 \neq i_3$ and $i_1 \neq i_4$) - expect. 0, | (i_1, i_2, i_1, i_4) ($i_1 \neq i_2$ and $i_1 \neq i_4$) - (p1),
(i_1, i_2, i_3, i_1) ($i_1 \neq i_2$ and $i_1 \neq i_3$) - expect. 0, | (i_1, i_2, i_2, i_4) ($i_2 \neq i_1$ and $i_2 \neq i_4$) - expect. 0,
(i_1, i_2, i_3, i_2) ($i_2 \neq i_1$ and $i_2 \neq i_3$) - (p2) relevant case ($i_1 = i_3$), see (p1),
(i_1, i_2, i_3, i_3) ($i_3 \neq i_1$ and $i_3 \neq i_2$) - expect. 0.
Three of the indices are equal and different from the fourth:
(i_1, i_2, i_2, i_2) ($i_1 \neq i_2$) - expect. 0, | (i_1, i_2, i_1, i_1) ($i_1 \neq i_2$) - (p3),
(i_1, i_1, i_3, i_1) ($i_1 \neq i_3$) - expect. 0, | (i_1, i_1, i_1, i_4) ($i_1 \neq i_4$) - (p4), same as (p3).
All the four indices are equal:
(i_1, i_1, i_1, i_1) - (p5).

Only the cases (p1), (p2), (p3), (p4) and (p5) have terms with non zero expectation. We now bound these expectations.

Case (p1):

$$\left|E\left(r_{2i_1}Z'_{i_1}V_{2,\delta}^{-1}Z_{i_2}x_{i_2h}\cdot r_{2i_3}Z'_{i_3}V_{2,\delta}^{-1}Z_{i_4}x_{i_4h}\right)\right| = \left|E\left(r_{2i_1}Z'_{i_1}V_{2,\delta}^{-1}Z_{i_2}x_{i_2h}\cdot r_{2i_1}Z'_{i_1}V_{2,\delta}^{-1}Z_{i_4}x_{i_4h}\right)\right|$$
$$= \left|E\left(x_{i_2h}Z'_{i_2}V_{2,\delta}^{-1}E\left[r_{2i_1}^2Z_{i_1}Z'_{i_1}\right]V_{2,\delta}^{-1}Z_{i_4}x_{i_4h}\right)\right|.$$

If $i_2\neq i_4$, by the independent sample assumption and the upper bounds on the eigenvalues of the matrices involved, for some $C>0$, this quantity is bounded by :

$$CE(x_{i_2h}Z'_{i_2})E(x_{i_2h}Z_{i_2}) = Cn^{-2\delta}a'_ka_k = O(n^{-2\delta}).$$

If $i_2=i_4$, by the same arguments, this quantity is bounded by:

$$CE(x_{i_2h}^2Z'_{i_2}Z_{i_2}) = C\cdot trace(E(x_{i_2h}^2Z_{i_2}Z'_{i_2})) = O(k).$$

It follows that

$$\left|\sum_{\text{indices in (p1)}}E\left(r_{2i_1}Z'_{i_1}V_{2,\delta}^{-1}Z_{i_2}x_{i_2h}\cdot r_{2i_3}Z'_{i_3}V_{2,\delta}^{-1}Z_{i_4}x_{i_4h}\right)\right| \leq n^3O(n^{-2\delta})+n^2O(k) = O(n^{3-2\delta}\vee n^2k).$$

Case (p2): This case corresponds to $i_2=i_4$, in (p1) and we claim that:

$$\left|\sum_{\text{indices in (p2)}}E\left(r_{2i_1}Z'_{i_1}V_{2,\delta}^{-1}Z_{i_2}x_{i_2h}\cdot r_{2i_3}Z'_{i_3}V_{2,\delta}^{-1}Z_{i_4}x_{i_4h}\right)\right| = O(n^2k).$$

Case (p3):

$$\left|E\left(r_{2i_1}Z'_{i_1}V_{2,\delta}^{-1}Z_{i_2}x_{i_2h}\cdot r_{2i_3}Z'_{i_3}V_{2,\delta}^{-1}Z_{i_4}x_{i_4h}\right)\right| = \left|E\left(r_{2i_1}Z'_{i_1}V_{2,\delta}^{-1}Z_{i_2}x_{i_2h}\cdot r_{2i_1}Z'_{i_1}V_{2,\delta}^{-1}Z_{i_1}x_{i_1h}\right)\right|$$
$$= n^{-\delta}\left|E\left(r_{2i_1}Z'_{i_1}V_{2,\delta}^{-1}a_k\cdot r_{2i_1}Z'_{i_1}V_{2,\delta}^{-1}Z_{i_1}x_{i_1h}\right)\right| \leq Cn^{-\delta}E\left(r_{2i_1}^2\cdot|x_{i_1h}|\cdot\|Z_{i_1}\|_2^3\right)$$
$$\leq Cn^{-\delta}\left(E(r_{2i_1}^{16/5}|x_{i_1h}|^{8/5})\right)^{5/8}\left(E(\|Z_{i_1}\|_2^8)\right)^{3/8} \leq Cn^{-\delta}k^{3/2}\left(k^{-1}\sum_{h=1}^kE(\tilde{Z}_{ih}^8)\right)^{3/8} = O(n^{-\delta}k^{3/2}),$$

with $\tilde{Z}_i = V_z^{-1/2}(Z_i-\mu_z)$. We use in the process the Holder's inequality and the Jensen's inequality. The constant C may differ from row to row. Thus,

$$\left|\sum_{\text{indices in (p3)}}E\left(r_{2i_1}Z'_{i_1}V_{2,\delta}^{-1}Z_{i_2}x_{i_2h}\cdot r_{2i_3}Z'_{i_3}V_{2,\delta}^{-1}Z_{i_4}x_{i_4h}\right)\right| = O(n^{2-\delta}k^{3/2}).$$

Case (p4): Same magnitude as (p3).

Case (p5):

$$\left|E\left(r_{2i_1}Z_{i_1}'V_{2,\delta}^{-1}Z_{i_2}x_{i_2h}\cdot r_{2i_3}Z_{i_3}'V_{2,\delta}^{-1}Z_{i_4}x_{i_4h}\right)\right| = \left|E\left(r_{2i}^2x_{ih}^2Z_i'V_{2,\delta}^{-1}Z_i\cdot Z_i'V_{2,\delta}^{-1}Z_ix_{ih}\right)\right|$$
$$\leq CE\left(r_{2i}^2x_{ih}^2\|Z_i\|_2^4\right) \leq C[E(r_{2i}^4x_{ih}^4)]^{1/2}[E(\|Z_i\|_2^8)]^{1/2} \leq Ck^2\left(k^{-1}\sum_{h=1}^k E(Z_{ih}^8)\right)^{1/2} = O(k^2).$$

Thus,

$$\left|\sum_{\text{indices in (p5)}} E\left(r_{2i_1}Z_{i_1}'V_{2,\delta}^{-1}Z_{i_2}x_{i_2h}\cdot r_{2i_3}Z_{i_3}'V_{2,\delta}^{-1}Z_{i_4}x_{i_4h}\right)\right| = O(nk^2).$$

Then, combining the contributions from (p1) to (p5), we claim that

$$E(C_{n,h}^2) = O(n^{1-2\delta}\vee k) + O(k) + O(n^{-2-\delta}k^{3/2}) + O(n^{-1}k^2) = O(k).$$

It results that $C_{n,h} = O_P(\sqrt{k})$ and, we deduce from (A.6) that $B_n'V^{-1}A_n = O_P(1)$. Hence,

$$\frac{B_n'V_{2,\delta}^{-1}A_n}{\sqrt{2k}} = O_P(k^{-1/2}). \tag{A.7}$$

Using (A.3), (A.4), (A.5) and (A.7), we conclude that

$$S_{n,k} = \frac{J_{n,k}-k}{\sqrt{2k}} = \frac{A_n'V_{2,\delta}^{-1}A_n - k}{\sqrt{2k}} + o_P(1)$$

and the result follows. □

Proof of Theorem 4.1: We have

$$\hat{\theta} = \theta_0 + \left(\tilde{\mu}_{zx}'\tilde{V}^{-1}\tilde{\mu}_{zx}\right)^{-1}\tilde{\mu}_{zx}'\tilde{V}^{-1}\tilde{\mu}_{z\varepsilon}, \quad \text{and} \quad \hat{V} = \frac{1}{n}\sum_{i=1}^n \hat{\varepsilon}^2(Z_i-\bar{Z})(Z_i-\bar{Z})'.$$

(a) Consider the case $0<\delta<1/2$. From Propositions OA.2(a) and OA.3(a) and their proofs, we have:

$$\tilde{e} = O_P(n^{\delta}),\ \ \tilde{V}^{-1} = V_3(\tilde{e})^{-1} + O_P(n^{-3\delta}),\ \ \|\tilde{\mu}_{zx}\|_2 = O_P(n^{-\delta}),\ \text{ and }\ \|\tilde{\mu}_{z\varepsilon}\|_2 = O_P(1).$$

Hence,

$$\tilde{\mu}_{zx}'\tilde{V}_3^{-1}\tilde{\mu}_{zx} = \tilde{\mu}_{zx}V(\tilde{e})^{-1}\tilde{\mu}_{zx} + O_P(n^{-5\delta}).$$

Since $\tilde{\mu}_{zx} = n^{-\delta}a_k + O_P(\sqrt{k/n})$ and $\|\tilde{e}\|_2^{-2} = O_P(n^{-2\delta})$, we obtain

$$\begin{aligned}\tilde{\mu}_{zx}'V_3(\tilde{e})^{-1}\tilde{\mu}_{zx} &= n^{-2\delta}a_k'V_3(\tilde{e})^{-1}a_k + \|V_3(\tilde{e})^{-1}\|_2\cdot O_P(n^{-\delta-1/2}k^{1/2})\\ &= \|\tilde{e}\|_2^{-2}n^{-2\delta}a_k'V_3(\tilde{e}/\|\tilde{e}\|_2)^{-1}a_k + O_P(n^{-\delta-1/2}k^{1/2}\|\tilde{e}\|_2^{-2})\\ &= \|\tilde{e}\|_2^{-2}n^{-2\delta}a_k'V_3(\tilde{u})^{-1}a_k + O_P(n^{-3\delta-1/2}\sqrt{k}).\end{aligned}$$

Hence,

$$\tilde{\mu}_{zx}\tilde{V}^{-1}\tilde{\mu}_{zx} = n^{-2\delta}\|\tilde{e}\|_2^{-2}a_k'V_3(\tilde{u})^{-1}a_k + O_P(n^{-3\delta-1/2}\sqrt{k}) \vee (n^{-5\delta}) := n^{-2\delta}\|\tilde{e}\|_2^{-2}a_k'V_3(\tilde{u})^{-1}a_k + \mathcal{E}.$$

Note that:

$$n^{-2\delta}\cdot\|\tilde{e}\|_2^{-2}\cdot a_k'V_3(\tilde{u})^{-1}a_k = O_P(n^{-4\delta}), \quad \text{and} \quad n^{2\delta}\cdot|\tilde{e}\|_2^{2}\cdot\left(a_k'V_3(\tilde{u})^{-1}a_k\right)^{-1} = O_P(n^{4\delta})$$

so that

$$n^{2\delta}\cdot|\tilde{e}\|_2^{2}\cdot\left(a_k'V_3(\tilde{u})^{-1}a_k\right)^{-1}\cdot\mathcal{E} = O_P(n^{\delta-1/2}\sqrt{k}\vee n^{-\delta}) = o_P(1).$$

Thus, following similar lines as in the proof of Proposition OA.3, we have:

$$\left(\tilde{\mu}_{zx}\tilde{V}^{-1}\tilde{\mu}_{zx}\right)^{-1} = n^{2\delta}\cdot\|\tilde{e}\|_2^{2}\cdot\left(a_k'V_3(\tilde{u})^{-1}a_k\right)^{-1} + O_P(n^{5\delta-1/2}\sqrt{k}\vee n^{3\delta}). \tag{A.8}$$

We now expand $\tilde{\mu}_{zx}\tilde{V}^{-1}\tilde{\mu}_{z\varepsilon}$. Note that

$$\tilde{\mu}_{z\varepsilon} = c_z + n^{-1}\sum_{i=1}^{n}(\varepsilon_i(Z_i-\mu_z)-c_z) + O_P(n^{-1}\sqrt{k}) = O_P(\|c_z\|_2) + O_P(\sqrt{k/n}) + O_P(n^{-1}\sqrt{k}) = O_P(1).$$

With this, it is not hard to see that:

$$\tilde{\mu}_{zx}\tilde{V}^{-1}\tilde{\mu}_{z\varepsilon} = n^{-\delta}\cdot\|\tilde{e}\|_2^{-2}\cdot a_k'V_3(\tilde{u})^{-1}c_z + O_P(n^{-2\delta-1/2}\sqrt{k}\vee n^{-4\delta}). \tag{A.9}$$

Using (A.8) and (A.9), we get:

$$\left(\tilde{\mu}_{zx}\tilde{V}^{-1}\tilde{\mu}_{zx}\right)^{-1}\cdot\tilde{\mu}_{zx}\tilde{V}^{-1}\tilde{\mu}_{z\varepsilon} = n^{\delta}\left(a_k'V_3(\tilde{u})^{-1}a_k\right)^{-1}a_k'V_3(\tilde{u})^{-1}c_z + O_P\left(n^{2\delta-1/2}\sqrt{k}\vee 1\right). \tag{A.10}$$

The result about $\hat{V}$ is derived along similar lines as in the proof of Proposition OA.3(a). We obtain that $\hat{e} = O_P(n^{\delta})$. The fact that $\liminf_k \inf_{\{u:\|u\|_2=1\}}\|a_k'V_3(u)^{-1}c_z\|_2 > 0$ and the eigenvalues of $(a_k'V_3(\tilde{u})^{-1}a_k)^{-1}$ are away from zero and from above as the sample size grows ensures that $\|\hat{e}\|_2^{-1} = O_P(n^{-\delta})$. The steps of the proof follows readily the same lines.

(b) Consider the case $\delta = 0$. We have: $\tilde{\mu}_{zx} = a_k + O_P(\sqrt{k/n})$. From Proposition OA.3(b), we can claim that $\tilde{\mu}_{zx}'\tilde{V}^{-1}\tilde{\mu}_{zx} = a_k'V_{3,0}^{-1}a_k + O_P(1/\sqrt{k})$ so that:

$$\left(\tilde{\mu}_{zx}'\tilde{V}^{-1}\tilde{\mu}_{zx}\right)^{-1} = \left(a_k'V_{3,0}^{-1}a_k\right)^{-1} + O_P(1/\sqrt{k}).$$

Also, it is not hard to see that

$$\tilde{\mu}_{zx}'\tilde{V}^{-1}\tilde{\mu}_{z\varepsilon} = a_k'V_{3,0}^{-1}c_z + O_P(1/\sqrt{k})$$

so that

$$\hat{\theta}-\theta_0 = \left(a_k'V_{3,0}^{-1}a_k\right)^{-1}a_k'V_{3,0}^{-1}c_z + O_P(1/\sqrt{k}).$$

The proof of the results on $\hat{V}$ follows the same lines as in that of Proposition OA.3(b).

(c) Consider the case $\delta \geq 1/2$. In this case,

$$\tilde{\mu}_{zx} = \bar{D}/\sqrt{n} + n^{-\delta} a_k + O_P(\sqrt{k}/n) = \bar{D}/\sqrt{n} + O_P(1/\sqrt{n}),$$

with $\|\bar{D}\|_2 = O_P(\sqrt{k})$. We have:

$$\tilde{\mu}_{zx}' \tilde{V}^{-1} \tilde{\mu}_{zx} = \tilde{\mu}_{zx}' V_3(\tilde{e})^{-1} \tilde{\mu}_{zx} + O_P(k^4/n^{5/2}).$$

Noting that $\|V_3(\tilde{e})^{-1}\|_2 = \|\tilde{e}\|_2^{-2} \|V_3(\tilde{u})^{-1}\|_2 = O_P(k^2/n)$, we obtain

$$\tilde{\mu}_{zx}' \tilde{V}^{-1} \tilde{\mu}_{zx} = \bar{D}' V_3(\tilde{e})^{-1} \bar{D}/n + O_P(k^{5/2}/n^2).$$

Note that $\|\bar{D}' V_3(\tilde{e})^{-1} \bar{D}/n\|_2 = O_P(k^3/n^2)$ and, under the conditions of the theorem, $\bar{D}' V_3(\tilde{e})^{-1} \bar{D}$ properly scaled is non-singular with probability approaching 1.
Hence, $\| \left(\bar{D}' V_3(\tilde{e})^{-1} \bar{D}/n\right)^{-1} \|_2 = O_P(n^2/k^3)$. We obtain the following inverse along similar lines as in the proof of Proposition OA.3. We get:

$$\begin{aligned}\left(\tilde{\mu}_{zx}' \tilde{V}^{-1} \tilde{\mu}_{zx}\right)^{-1} &= \left(\bar{D}' V_3(\tilde{e})^{-1} \bar{D}/n\right)^{-1} + O_P\left(n^2/k^{7/2}\right) \\ &= \frac{n}{k} \|\tilde{e}\|_2^2 \left(k^{-1/2} \bar{D}' V_3(\tilde{u})^{-1} k^{-1/2} \bar{D}\right)^{-1} + O_P\left(n^2/k^{7/2}\right).\end{aligned}$$

Also, we have:

$$\tilde{\mu}_{zx}' \tilde{V}^{-1} \tilde{\mu}_{z\varepsilon} = \tilde{\mu}_{zx}' V_3(\tilde{e})^{-1} \tilde{\mu}_{z\varepsilon} + O_P(k^{7/2}/n^2).$$

Straightforward derivations yield:

$$\tilde{\mu}_{zx}' \tilde{V}^{-1} \tilde{\mu}_{z\varepsilon} = \frac{\|\tilde{e}\|_2^{-2}}{\sqrt{n}} \left(\bar{D}' V_3(\tilde{u})^{-1} c_z + n^{-\delta+1/2} a_k' V_3(\tilde{u})^{-1} c_z\right) + O_P(k^{7/2}/n^2).$$

It follows that:

$$\hat{\theta} - \theta_0 = \frac{\sqrt{n}}{k} \cdot \left(k^{-1/2} \bar{D}' V_3(\tilde{u})^{-1} k^{-1/2} \bar{D}\right)^{-1} \left(\bar{D}' V_3(\tilde{u})^{-1} c_z + n^{-\delta+1/2} a_k' V_3(\tilde{u})^{-1} c_z\right) + O_P(\sqrt{n}/k^{3/2}).$$

The proof of the results on $\hat{V}$ follows the same lines as in that of Proposition OA.3(c). Assumption 6(b) ensures that $\hat{e} := \hat{\theta} - \theta_0 = O_P(\sqrt{n}/k)$ and this is sufficient to claim, as in the proof of Proposition OA.3(a), that $\hat{V} = V_3(\hat{e}) + O_P(\sqrt{n}/k)$. Under Assumption 6(c), regardless of $\delta = 1/2$ or $\delta > 1/2$, we can claim that $\|\hat{e}\|^{-1} = O_P(k/\sqrt{n})$ and this is sufficient to claim the same conclusion as Proposition OA.3(c) regarding $\hat{V}^{-1}$. □

Proof of Theorem 4.2: Recall that $J_{n,k} = n\left(\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta}\right)' \hat{V}^{-1}\left(\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta}\right)$ and $S_{n,k} = (J_{n,k} - k)/\sqrt{2k}$. Using the expression of $\hat{\theta} - \theta_0$ given by Theorem 4.1, we have:

$$\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta} = \bar{\mu}_{z\varepsilon} - \bar{\mu}_{zx}(\hat{\theta} - \theta_0) - (\bar{Z} - \mu_z)(\bar{x} - \mu_x)'(\hat{\theta} - \theta_0). \tag{A.11}$$

(a) $0 < \delta < 1/2$. We have:

$$\begin{aligned}
&\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta} = c_z - a_k(a_k' V_3(\tilde{u})^1 a_k)^{-1} a_k' V_3(\tilde{u})^{-1} c_z + O_P(n^{\delta-1/2}\sqrt{k} \vee n^{-\delta}) \\
&= V_3(\tilde{u})^{1/2}\left(I_k - V_3(\tilde{u})^{-1/2} a_k (a_k' V_3(\tilde{u})^{-1} a_k)^{-1} a_k' V_3(\tilde{u})^{-1/2}\right) V_3(\tilde{u})^{-1/2} c_z + O_P(n^{\delta-1/2}\sqrt{k} \vee n^{-\delta}) \\
&\qquad := V_3(\tilde{u})^{1/2}(I_k - P_\delta) V_3(\tilde{u})^{-1/2} c_z + O_P(n^{\delta-1/2}\sqrt{k} \vee n^{-\delta}) = O_P(1).
\end{aligned}$$

From Theorem 4.1(a), we can claim that

$$\begin{aligned}
&\left(\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta}\right)' \hat{V}^{-1}\left(\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta}\right) = \left(\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta}\right)' V_3(\hat{e})^{-1}\left(\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta}\right) + O_P(n^{-3\delta}) \\
&= \|\hat{e}\|_2^{-2} c_z' V_3(\tilde{u})^{-1/2}(I_k - P_\delta) V_3(\tilde{u})^{1/2} V_3(\hat{u})^{-1} V_3(\tilde{u})^{1/2}(I_k - P_\delta) V_3(\tilde{u})^{-1/2} c_z + O_P(n^{-\delta-1/2}\sqrt{k} \vee n^{-3\delta}).
\end{aligned}$$

Note that under Assumption 4(a), $\|V_3(\hat{u})\|_2 \le C$ for some $C > 0$. Also, thanks to Assumption 5, $\lambda_{\min}(V_3(\tilde{u})) \ge \underline{\lambda}$. [In the next lines, we use C as a generic positive constant with value that may change with the context.] Thus, for some $C > 0$, we claim that

$$\begin{aligned}
&\left(\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta}\right)' \hat{V}^{-1}\left(\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta}\right) \\
&\qquad \ge C\|\hat{e}\|_2^{-2} c_z' V_3(\tilde{u})^{-1/2}(I_k - P_\delta) V_3(\tilde{u})^{-1/2} c_z + O_P(n^{-\delta-1/2}\sqrt{k} \vee n^{-3\delta}).
\end{aligned} \tag{A.12}$$

Besides, note that

$$\|\hat{e}\|_2^2 = \hat{e}'\hat{e} = n^{2\delta} \cdot c_z' V_3(\tilde{u})^{-1} a_k (a_k' V_3(\tilde{u})^{-1} a_k)^{-2} a_k' V_3(\tilde{u})^{-1} c_z + O_P(n^{3\delta-1/2}\sqrt{k} \vee n^{\delta}).$$

But, for any k large enough,

$$\lambda_{\max}((a_k' V_3(\tilde{u})^{-1} a_k)^{-1}) \le \frac{\lambda_{\max}(V_3(\tilde{u}))}{\lambda_{\min}(a_k' a_k)} \le C,$$

for some positive constant C, where we use Assumption 1(b) and the fact that $\|V_3(\tilde{u})\|_2$ is bounded under the maintained assumptions. It follows that:

$$\begin{aligned}
\|\hat{e}\|_2^2 &\le n^{2\delta} \cdot C \cdot c_z' V_3(\tilde{u})^{-1/2} P_\delta V_3(\tilde{u})^{-1/2} c_z + O_P(n^{3\delta-1/2}\sqrt{k} \vee n^{\delta}) \\
&\qquad \le n^{2\delta} \cdot C + O_P(n^{3\delta-1/2}\sqrt{k} \vee n^{\delta}),
\end{aligned}$$

(for a different value of $C > 0$), where we use the fact that $\lambda_{\min}(V_3(\tilde{u})) \ge \underline{\lambda}$ and $c_z' c_z = O(1)$. Therefore,

$$\|\hat{e}\|_2^{-2} \ge n^{-2\delta} \cdot C \cdot \left(1 + O_P(n^{\delta-1/2}\sqrt{k} \vee n^{-\delta})\right).$$

From (A.12), we can state that:

$$\left(\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta}\right)' \hat{V}^{-1} \left(\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta}\right)$$
$$\geq C \cdot n^{-2\delta} \cdot c_z' V_3(\tilde{u})^{-1/2}(I_k - P_\delta)V_3(\tilde{u})^{-1/2}c_z + O_P(n^{-\delta-1/2}\sqrt{k} \vee n^{-3\delta}) \quad \text{(A.13)}$$

and

$$J_{n,k} \geq n^{1-2\delta} \cdot C \cdot \Delta_k + O_P(n^{1/2-\delta}\sqrt{k} \vee n^{1-3\delta})$$

and the statement in part (a) follows readily.

(b) $\delta = 0$. Using Theorem 4.1(b), its is not hard to see that

$$\bar{\mu}_{zx}(\hat{\theta} - \theta_0) = a_k(a_k' V_{3,0}^{-1} a_k)^{-1} a_k' V_{3,0}^{-1} c_z + O_P(1/\sqrt{k}).$$

Using (A.11), we have:

$$\begin{aligned}\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta} &= V_{3,0}^{1/2} \left(I_k - V_{3,0}^{-1/2} a_k(a_k' V_{3,0}^{-1} a_k)^{-1} a_k' V_{3,0}^{-1/2}\right) V_{3,0}^{-1/2} c_z + O_P(1/\sqrt{k}) \\ &= V_{3,0}^{1/2}(I_k - P_\delta)V_{3,0}^{-1/2} c_z + O_P(1/\sqrt{k}).\end{aligned}$$

It follows that

$$\begin{aligned}J_{n,k} &= n\left(c_z' V_{3,0}^{-1/2}(I_k - P_\delta)V_{3,0}^{1/2}\hat{V}^{-1}V_{3,0}^{1/2}(I_k - P_\delta)V_{3,0}^{-1/2}c_z + O_P(1/\sqrt{k})\right) \\ &= n\left(c_z' V_{3,0}^{-1/2}(I_k - P_\delta)V_{3,0}^{1/2}V_{3,1}^{-1}V_{3,0}^{1/2}(I_k - P_\delta)V_{3,0}^{-1/2}c_z + O_P(1/\sqrt{k})\right) \\ &\geq \bar{\lambda}^{-1}\underline{\lambda} \cdot n \cdot c_z' V_{3,0}^{-1/2}(I_k - P_\delta)V_{3,0}^{-1/2}c_z + O_P(n/\sqrt{k}).\end{aligned}$$

It follows that, if as k grows c_z does not lie in the column span of a_k, then $J_{n,k} \to \infty$ in probability as $n \to \infty$ and $S_{n,k} = (J_{n,k} - k)/\sqrt{2k} = O_P(n/\sqrt{k}) \to \infty$ in probability as $n \to \infty$.

(c) $\delta \geq 1/2$. Using Theorem 4.1(c), we can see that $\hat{\theta} - \theta_0 = O_P(\sqrt{n}/k)$. The fact that $\bar{\mu}_{z\varepsilon} = c_z + O_P(\sqrt{k/n})$ and $\bar{\mu}_{zx} = O_P(\sqrt{k/n})$, allows us to claim that

$$\tilde{\mu}_{zy} - \tilde{\mu}_{zy}\hat{\theta} = c_z + O_P(1/\sqrt{k}).$$

Again, using Theorem 4.1(c), we claim that:

$$\begin{aligned}J_{n,k} :=& n(\tilde{\mu}_{zy} - \tilde{\mu}_{zy}\hat{\theta})'\hat{V}^{-1}(\tilde{\mu}_{zy} - \tilde{\mu}_{zy}\hat{\theta}) \\ =& n(c_z + O_P(1/\sqrt{k}))'\left(V_3(\hat{e})^{-1} + O_P(k^3/n^{3/2})\right)(c_z + O_P(1/\sqrt{k})) \\ =& nc_z' V_3(\hat{e})^{-1}c_z + nc_z' V_3(\hat{e})^{-1} \cdot O_P(1/\sqrt{k}) + O_P(k^3/\sqrt{n}) \\ =& n\|\hat{e}\|_2^{-2} \cdot c_z' V_3(\hat{u})^{-1}c_z + O_P(n/\sqrt{k}) \cdot \|\hat{e}\|_2^{-2} + O_P(k^3/\sqrt{n}) \\ \geq& [\lambda_{\max}(V_3(\hat{u})]^{-1} n\|\hat{e}\|_2^{-2} \cdot \|c_z\|_2^2 + O_P(n/\sqrt{k}) \cdot \|\hat{e}\|_2^{-2} + O_P(k^3/\sqrt{n}) \\ \geq& C \cdot n\|\hat{e}\|_2^{-2}\left(\|c_z\|_2^2 + O_P(1/\sqrt{k})\right) + O_P(k^3/\sqrt{n}),\end{aligned}$$

with $\hat{u} = \hat{e}/\|\hat{e}\|_2$ and for some constant $C > 0$. The last inequality uses the fact that $\lambda_{\max}(V_3(u))$ is uniformly bound from above over $u : \|u\|_2 = 1$ and n.

Since $\|c_z\|_2^2$ is nondecreasing in n and is nonzero for n large enough, we claim that, with probability approaching 1,

$$J_{n,k} \geq (C/2) \cdot \|\hat{e}\|_2^{-2} \cdot \|c_z\|_2^2 + O_P(k^3/\sqrt{n}) = (C/2) \cdot k^2 \cdot \left(nk^{-2}\|\hat{e}\|_2^{-2}\right) \cdot \|c_z\|_2^2 + O_P(k^3/\sqrt{n}).$$

Letting $\pi_n := (C/2) \cdot \left(nk^{-2}\|\hat{e}\|_2^{-2}\right)$,

$$J_{n,k} \geq k^2\pi_n\|c_z\|_2^2 + O_P(k^3/\sqrt{n}).$$

Since $(k^2/n)\|\hat{e}\|_2^2 = O_P(1)$, we can claim by definition that $\lim_{\epsilon\downarrow 0}\sup_n P(\pi_n < \epsilon) = 0$ and this establishes the first part of the statement.

We can also claim that, with probability approaching one,

$$\begin{aligned} S_{n,k} \geq &(2k)^{-1/2}\left(k^2\pi_n\|c_z\|_2^2 - k + O_P(k^3/\sqrt{n})\right) = 2^{-1/2}k^{3/2}\pi_n\left(\|c_z\|_2^2 + O_P(1/k)\right) + O_P(k^{5/2}/\sqrt{n}) \\ \geq &(1/2\sqrt{2})k^{3/2}\pi_n\|c_z\|_2^2 + O_P(k^{5/2}/\sqrt{n}) \end{aligned}$$

and this concludes the proof. □

Online Appendix for

“A uniformly valid test for instrument exogeneity”

Prosper Dovonon and Nikolay Gospodinov

This Online Appendix provides some additional theoretical and simulation results and is organized as follows. Section OA.1 presents the proof of Theorem 2.1 in the paper. Section OA.2 describes the asymptotic order of magnitude of the GMM estimator under the alternative hypothesis through Propositions OA.2 and OA.3 along with their respective proofs. Proposition OA.2 focuses on the case where the weighting matrix does not depend on parameter estimate while Proposition OA.3 focuses on GMM with the optimal weighting matrix. In Section OA.3, we provide two lemmas (Lemma OA.4 and Lemma OA.5) showing that Δ_k in Theorem 4.2 does not vanish as k grows. Section OA.4 derives the asymptotic distribution of the GMM estimator in the case where $0 \leq \delta < 1/2$. Section OA.5 considers a simple model with a single regressor and completely irrelevant instruments to provide further intuition and clarification of the surprising power properties of the test. Finally, Section OA.6 reports additional simulation results for different sample sizes and fat-tailed distributions.

OA.1 Proof of Theorem 2.1

Proof of Theorem 2.1: We have:

$$\tilde{\theta} = \theta_0 + \left(\tilde{\mu}'_{zx}\hat{W}\tilde{\mu}_{zx}\right)^{-1}\left(\tilde{\mu}'_{zx}\hat{W}\tilde{\mu}_{z\varepsilon}\right).$$

We derive this result in four steps. In Steps I and II, we derive the orders of magnitude of $\tilde{\mu}'_{zx}W\tilde{\mu}_{zx}$ and $\tilde{\mu}'_{zx}W\tilde{\mu}_{z\varepsilon}$. Step III deals with the magnitudes of the feasible quantities with W replaced by $\hat{W}$ in Steps I and II. Finally, we derive in Step IV the orders of the inverse of $\tilde{\mu}'_{zx}\hat{W}\tilde{\mu}_{zx}$ and conclude the proof.

I. Consider first $\tilde{\mu}'_{zx}W\tilde{\mu}_{zx}$. We have:

$$\tilde{\mu}'_{zx}W\tilde{\mu}_{zx} = \bar{\mu}'_{zx}W\bar{\mu}_{zx} - \bar{\mu}'_{zx}W(\bar{Z}-\mu_z)(\bar{x}-\mu_x)' - (\bar{x}-\mu_x)(\bar{Z}-\mu_z)'W\bar{\mu}_{zx} + (\bar{x}-\mu_x)(\bar{Z}-\mu_z)'W(\bar{Z}-\mu_z)(\bar{x}-\mu_x).$$

We observe that:

$$\bar{\mu}_{zx} = n^{-\delta}a_k + O_P(\sqrt{k/n}) = O_P(n^{-\delta} \vee \sqrt{k/n}) \quad \text{(OA.1)}$$

and, using the fact that: $\bar{x}-\mu_x = O_P(n^{-1/2})$, and $\bar{Z}-\mu_z = O_P(\sqrt{k/n})$, we obtain:

$$\tilde{\mu}'_{zx}W\tilde{\mu}_{zx} = \bar{\mu}'_{zx}W\bar{\mu}_{zx} + O_P\left(\frac{\sqrt{k}}{n^{1+\delta}} \vee \frac{k}{n\sqrt{n}}\right). \quad \text{(OA.2)}$$

Now, we consider $\bar{\mu}'_{zx}W\bar{\mu}_{zx}$. We have:

$$\bar{\mu}'_{zx}W\bar{\mu}_{zx} = \frac{1}{n^2}\sum_{i=1}^{n}(Z_i-\mu_z)'W(Z_i-\mu_z)\cdot(x_i-\mu_x)(x_i-\mu_x)' + \frac{1}{n^2}\sum_{i\neq j}(Z_i-\mu_z)'W(Z_j-\mu_z)\cdot(x_i-\mu_x)(x_j-\mu_x)' := (1)+(2). \quad \text{(OA.3)}$$

I.1. Let us consider (2). We have:

$$E(2) = \frac{1}{n^2}\sum_{i\neq j}E\left((x_i-\mu_x)(Z_i-\mu_z)'\right)WE\left((Z_i-\mu_z)(x_i-\mu_x)'\right) = \left(1-\frac{1}{n}\right)n^{-2\delta}a'_kWa_k.$$

Now, consider the (h,h')-component of $(2)-E(2) := (2)_{h,h'} - E[(2)_{h,h'}]$; for $h,h' = 1,\dots,p$. We obtain the order of magnitude of this quantity by deriving its mean-square. We have:

$$E\left((2)_{h,h'} - E[(2)_{h,h'}]\right)^2 =$$

$$\frac{1}{n^4}E\left(\sum_{i\neq j}(x_{ih}-\mu_{xh})(x_{ih'}-\mu_{xh'})(Z_i-\mu_z)'W(Z_j-\mu_z) - n^{-2\delta}a'_{kh}Wa_{kh'}\right)^2 := \frac{1}{n^4}E\left(\sum_{i\neq j}b_{i,j}\right)^2,$$

where a_{kh} is the h-th column of a_k. It is not hard to see that:

$$E\left(\sum_{i\neq j} b_{i,j}\right)^2 = \sum_{i\neq j} E(b_{i,j}^2) + \sum_{i\neq j, i\neq j', j\neq j'} E(b_{i,j}\cdot b_{i,j'}) + \sum_{i\neq j, i'\neq j, i\neq i'} E(b_{i,j}\cdot b_{i',j}).$$

In the next expansions, we will use at times the notation $\tilde{x}_i := x_i - \mu_x$, $\tilde{x}_{ih} := x_{ih} - \mu_{xh}$.

$$\begin{aligned}
E(b_{i,j}^2) &\leq E\left(\tilde{x}_{ih}\tilde{x}_{ih'}(Z_i-\mu_z)'W(Z_j-\mu_z)\right)^2 \\
&= E\left(\tilde{x}_{ih}(Z_i-\mu_z)'WE\left[\tilde{x}_{jh'}^2(Z_j-\mu_z)(Z_j-\mu_z)'\right]W\tilde{x}_{ih}(Z_i-\mu_z)\right) \\
&\leq \bar{\lambda}^3 trace[E\left[\tilde{x}_{ih}^2(Z_i-\mu_z)(Z_i-\mu_z)'\right]] \leq \bar{\lambda}^4 k
\end{aligned}$$

and

$$\begin{aligned}
&E(b_{i,j}\cdot b_{i,j'}) \\
&= E\left(\tilde{x}_{ih}^2\tilde{x}_{jh'}\tilde{x}_{j'h'}\cdot(Z_i-\mu_z)'W(Z_j-\mu_z)\cdot(Z_i-\mu_z)'W(Z_{j'}-\mu_z)\right) - n^{-4\delta}\left(a_{kh}'Wa_{kh'}\right)^2 \\
&= E\left(\tilde{x}_{jh'}(Z_j-\mu_z)'\right)WE\left(\tilde{x}_{ih}^2(Z_i-\mu_z)(Z_i-\mu_z)'\right)WE\left(\tilde{x}_{j'h'}(Z_{j'}-\mu_z)\right) + O(n^{-4\delta}) \\
&= n^{-\delta}a_{kh}'WE\left(\tilde{x}_{ih}^2(Z_i-\mu_z)(Z_i-\mu_z)'\right)Wn^{-\delta}a_{kh'} + O(n^{-4\delta}) = O(n^{-2\delta}).
\end{aligned}$$

It follows that:

$$\frac{1}{n^4}E\left(\sum_{i\neq j} b_{i,j}\right)^2 = \frac{n(n-1)}{n^4}O(k) + \frac{n(n-1)(n-2)}{n^4}O(n^{-2\delta}) = O\left(\frac{k}{n^2}\right) + O\left(\frac{1}{n^{1+2\delta}}\right).$$

Thus, for all $h, h' = 1, \dots, p$,

$$(2)_{hh'} = E(2)_{hh'} + O_P\left(\frac{\sqrt{k}}{n}\vee\frac{1}{n^{\delta+1/2}}\right) = n^{-2\delta}a_{kh}'Wa_{kh'} + O_P\left(\frac{\sqrt{k}}{n}\vee\frac{1}{n^{\delta+1/2}}\right).$$

As a result,

$$(2) = n^{-2\delta}a_k'Wa_k + O_P\left(\frac{\sqrt{k}}{n}\vee\frac{1}{n^{\delta+1/2}}\right). \tag{OA.4}$$

I.2. Let us consider (1) as defined by (OA.3). We have:

$$n\cdot(1) = \frac{1}{n}\sum_{i=1}^{n}(Z_i-\mu_z)'W(Z_i-\mu_z)\cdot\tilde{x}_i\tilde{x}_i'.$$

For $h, h' = 1, \dots, p$,

$$n\cdot(1)_{hh'} = \frac{1}{n}\sum_{i=1}^{n}\tilde{x}_{ih}\tilde{x}_{ih'}\cdot(Z_i-\mu_z)'W(Z_i-\mu_z).$$

Note that

$$E(n \cdot (1)_{hh'}) = E\tilde{x}_{ih}\tilde{x}_{ih'} \cdot (Z_i - \mu_z)'W(Z_i - \mu_z) := k \cdot \mathbb{V}_{1k,hh'}.$$

The mean square error is given by:

$$\begin{aligned}
E\left(n \cdot (1)_{hh'} - k \cdot \mathbb{V}_{1k,hh'}\right)^2 &= E\left(\frac{1}{n}\sum_{i=1}^{n}[\tilde{x}_{ih}\tilde{x}_{ih'}(Z_i - \mu_z)'W(Z_i - \mu_z) - k \cdot \mathbb{V}_{1k,hh'}]\right)^2 \\
&:= E\left(\frac{1}{n}\sum_{i=1}^{n} b_i\right)^2 = \frac{1}{n^2}\sum_{i=1}^{n} E(b_i^2) = n^{-1}E(b_i^2) \\
&\leq n^{-1}E\left(\tilde{x}_{ih}^2\tilde{x}_{ih'}^2[(Z_i - \mu_z)'W(Z_i - \mu_z)]^2\right) = n^{-1}E\left(\tilde{x}_{ih}^2\tilde{x}_{ih'}^2[\tilde{Z}_i'V_z^{1/2}WV_z^{1/2}\tilde{Z}_i]^2\right) \\
&\leq n^{-1}\bar{\lambda}^4 E\left(\tilde{x}_{ih}^2\tilde{x}_{ih'}^2[\tilde{Z}_i'\tilde{Z}_i]^2\right) \leq n^{-1}\bar{\lambda}^4 k^2 C\left(Ek^{-1}\sum_{h=1}^{k}\tilde{Z}_{ih}^8\right)^{1/2} = O(k^2/n),
\end{aligned}$$

where C is a positive constant. The last inequality is obtained by applying the Cauchy-Schwarz inequality and then the Jensen's inequality to $[k^{-1}\tilde{Z}_i'\tilde{Z}_i]^4$ and then we use Assumptions 1(b) and 2(d) to conclude.

It follows that

$$n \cdot (1)_{hh'} - k \cdot \mathbb{V}_{1k,hh'} = O_P(k/\sqrt{n}),$$

that is:

$$(1) = \frac{1}{n^2}\sum_{i=1}^{n}(Z_i - \mu_z)'W(Z_i - \mu_z) \cdot \tilde{x}_i\tilde{x}_i' = \frac{k}{n}\mathbb{V}_{1k} + O_P\left(\frac{k}{n\sqrt{n}}\right). \tag{OA.5}$$

Using equations (OA.2), (OA.3), (OA.4) and (OA.5), we obtain:

$$\bar{\mu}_{zx}'W\bar{\mu}_{zx} = \frac{k}{n}\mathbb{V}_{1k} + n^{-2\delta}a_k'Wa_k + O_P\left(\frac{\sqrt{k}}{n} \vee \frac{1}{n^{\delta+1/2}}\right) \tag{OA.6}$$

and

$$\tilde{\mu}_{zx}'W\tilde{\mu}_{zx} = \frac{k}{n}\mathbb{V}_{1k} + n^{-2\delta}a_k'Wa_k + O_P\left(\frac{\sqrt{k}}{n} \vee \frac{1}{n^{\delta+1/2}}\right). \tag{OA.7}$$

It follows that:

- If $\delta \geq 1/2$,

$$\tilde{\mu}_{zx}'W\tilde{\mu}_{zx} = \frac{k}{n}\mathbb{V}_{1k} + O_P\left(\frac{\sqrt{k}}{n}\right).$$

- If $0 \leq \delta < 1/2$,

  – If $k \ll n^{1-2\delta}$,

$$\tilde{\mu}_{zx}'W\tilde{\mu}_{zx} = n^{-2\delta}a_k'Wa_k + O_P\left(\frac{k}{n} \vee \frac{1}{n^{\delta+1/2}}\right).$$

– If $\quad k \sim n^{1-2\delta}$,

$$\tilde{\mu}_{zx}' W \tilde{\mu}_{zx} = \frac{1}{n^{2\delta}} \left(\mathbb{V}_{1k} + a_k' W a_k \right) + O_P \left(\frac{1}{n^{\delta+1/2}} \right).$$

– If $\quad k \gg n^{1-2\delta}$,

$$\tilde{\mu}_{zx}' W \tilde{\mu}_{zx} = \frac{k}{n} \mathbb{V}_{1k} + O_P \left(\frac{\sqrt{k}}{n} \vee \frac{1}{n^{2\delta}} \right).$$

II. Consider second $\tilde{\mu}_{zx}' W \tilde{\mu}_{z\varepsilon}$. We have:

$$\tilde{\mu}_{zx} = \bar{\mu}_{zx} - (\bar{Z} - \mu_z)(\bar{x} - \mu_x)'; \quad \tilde{\mu}_{z\varepsilon} = \bar{\mu}_{z\varepsilon} - (\bar{Z} - \mu_z)\bar{\varepsilon}.$$

Hence,

$$\tilde{\mu}_{zx}' W \tilde{\mu}_{z\varepsilon} = \bar{\mu}_{zx}' W \bar{\mu}_{z\varepsilon} - \bar{\mu}_{zx}' W (\bar{Z} - \mu_z)\bar{\varepsilon} - (\bar{x} - \mu_x)(\bar{Z} - \mu_z)' W \bar{\mu}_{z\varepsilon} + (\bar{x} - \mu_x)(\bar{Z} - \mu_z)' W (\bar{Z} - \mu_z)\bar{\varepsilon}.$$

Using (OA.3), the fact that $\bar{\mu}_{z\varepsilon} = O_P(\sqrt{k/n})$, along with the orders of magnitude of $\bar{Z} - \mu_z$ and $\bar{x} - \mu_x$, we can claim that:

$$\tilde{\mu}_{zx}' W \tilde{\mu}_{z\varepsilon} = \bar{\mu}_{zx}' W \bar{\mu}_{z\varepsilon} + O_P \left(\frac{\sqrt{k}}{n^{1+\delta}} \vee \frac{k}{n\sqrt{n}} \right). \tag{OA.8}$$

Consider now: $\bar{\mu}_{zx}' W \bar{\mu}_{z\varepsilon}$. We have:

$$\bar{\mu}_{zx}' W \bar{\mu}_{z\varepsilon} = \frac{1}{n^2} \sum_{i=1}^{n} \tilde{x}_i \varepsilon_i (Z_i - \mu_z)' W (Z_i - \mu_z) + \frac{1}{n^2} \sum_{i \neq j} \tilde{x}_i (Z_i - \mu_z)' W (Z_j - \mu_z) \varepsilon_j := (1) + (2). \tag{OA.9}$$

II.1 Consider (2). It is not hard to see that, under the null, $E((2)) = 0$. The mean square of the h-th component of (2) is given by:

$$E((2)_h)^2 = \frac{1}{n^4} E \left(\sum_{i \neq j} \tilde{x}_{ih} (Z_i - \mu_z)' W (Z_j - \mu_z) \varepsilon_j \right)^2 := \frac{1}{n^4} E \left(\sum_{i \neq j} b_{i,j} \right)^2$$

$$= \frac{1}{n^4} \left(\sum_{i \neq j} E(b_{i,j}^2) + \sum_{i \neq j} E(b_{i,j} b_{j,i}) + \sum_{i \neq j, i \neq i', j \neq i'} E(b_{i,j} b_{i',j}). \right)$$

$$\begin{aligned} E(b_{i,j}^2) &= E \left(\tilde{x}_{ih}^2 \varepsilon_j^2 (Z_i - \mu_z)' W (Z_j - \mu_z)(Z_j - \mu_z)' W (Z_i - \mu_z) \right) \\ &= E \left[\tilde{x}_{ih}^2 (Z_i - \mu_z)' W \cdot E(\varepsilon_j^2 (Z_j - \mu_z)(Z_j - \mu_z)') \cdot W (Z_i - \mu_z) \right] \\ &\leq \bar{\lambda}^3 E(\tilde{x}_{ih}^2 (Z_i - \mu_z)'(Z_i - \mu_z)) \leq \bar{\lambda}^4 \cdot k = O(k). \end{aligned}$$

By the Cauchy-Schwarz inequality, this bound also implies that: $|E(b_{i,j} \cdot b_{j,i})| = O(k)$.

$$\begin{aligned} E(b_{ij} \cdot b_{i'j}) &= E\left[\tilde{x}_{ih}(Z_i - \mu_z)'W(Z_j - \mu_z)\varepsilon_j \cdot \tilde{x}_{i'h}(Z_{i'} - \mu_z)'W(Z_j - \mu_z)\varepsilon_j\right] \\ &= E\left(\tilde{x}_{ih}(Z_i - \mu_z)'\right) W E\left(\varepsilon_j^2(Z_j - \mu_z)(Z_j - \mu_z)'\right) W E\left(\tilde{x}_{i'h}(Z_{i'} - \mu_z)\right) \\ &= n^{-\delta} a'_{kh} W E\left(\varepsilon_j^2 (Z_j - \mu_z)(Z_j - \mu_z)'\right) W n^{-\delta} a_{kh} \leq \bar{\lambda}^3 n^{-2\delta} a'_{kh} a_{kh} = O(n^{-2\delta}). \end{aligned}$$

Thus, $E((2)_h)^2 = O(k/n^2) + O(1/n^{1+2\delta})$ and it follows that:

$$(2) = O_P\left(\frac{\sqrt{k}}{n} \vee \frac{1}{n^{\delta+1/2}}\right). \tag{OA.10}$$

II.2 Consider now (1).

$$n \cdot (1) = n^{-1} \sum_{i=1}^{n} \tilde{x}_i \varepsilon_i (Z_i - \mu_z)'W(Z_i - \mu_z).$$

For $h = 1, \ldots, p$, consider the h-th component of $n \cdot (1)$, that is: $n \cdot (1)_h$. We have:

$$E\left(n \cdot (1)_h\right) = E\left(\tilde{x}_{ih}\varepsilon_i(Z_i - \mu_z)'W(Z_i - \mu_z)\right) = k \cdot \mathbb{C}_{1k,h},$$

where the last equality follows by definition of $\mathbb{C}_{1k}$. Also,

$$\begin{aligned} E\left(n \cdot (1)_h - k \cdot \mathbb{C}_{1k,h}\right)^2 &= \frac{1}{n} E\left(\tilde{x}_{ih}\varepsilon_i(Z_i - \mu_z)'W(Z_i - \mu_z) - k \cdot \mathbb{C}_{1k,h}\right)^2 \\ &\leq \frac{1}{n} E\left(\tilde{x}_{ih}\varepsilon_i(Z_i - \mu_z)'W(Z_i - \mu_z)\right)^2 \leq \frac{\bar{\lambda}^4}{n} E\left(\tilde{x}_{ih}^2 \varepsilon_i^2 (\tilde{Z}_i'\tilde{Z}_i)^2\right) \\ &\leq \frac{\bar{\lambda}^4}{n} \cdot C \cdot k^2 \cdot k^{-1} \sum_{l=1}^{k} E\tilde{Z}_{il}^8 = O(k^2/n) \end{aligned}$$

for some $C > 0$, where the last inequality follows from the Cauchy-Schwarz and the Jensen's inequalities. Therefore, we have: $(1)_h = (k/n)\mathbb{C}_{1k,h} + O_P(k/n\sqrt{n})$. That is:

$$(1) = \frac{k}{n}\mathbb{C}_{1k} + O_P\left(\frac{k}{n\sqrt{n}}\right). \tag{OA.11}$$

Hence,

$$\bar{\mu}'_{zx} W \bar{\mu}_{z\varepsilon} = (1) + (2) = \frac{k}{n}\mathbb{C}_{1k} + O_P\left(\frac{\sqrt{k}}{n} \vee \frac{1}{n^{\delta+1/2}}\right). \tag{OA.12}$$

Also, using (OA.8), (OA.10), and (OA.11), we obtain:

$$\tilde{\mu}'_{zx} W \tilde{\mu}_{z\varepsilon} = \frac{k}{n}\mathbb{C}_{1k} + O_P\left(\frac{\sqrt{k}}{n} \vee \frac{1}{n^{\delta+1/2}}\right). \tag{OA.13}$$

The following cases arise:

- If $\delta \geq 1/2$,
$$\tilde{\mu}'_{zx} W \tilde{\mu}_{z\varepsilon} = \frac{k}{n}\mathbb{C}_{1k} + O_P\left(\frac{\sqrt{k}}{n}\right).$$

- If $0 \leq \delta < 1/2$,
  - If $k \ll n^{1/2-\delta}$ or $k \sim n^{1/2-\delta}$,
$$\tilde{\mu}'_{zx} W \tilde{\mu}_{z\varepsilon} = O_P(n^{-1/2-\delta}).$$
  - If $n^{1/2-\delta} \ll k \ll n^{1-2\delta}$,
$$\tilde{\mu}'_{zx} W \tilde{\mu}_{z\varepsilon} = \frac{k}{n}\mathbb{C}_{1k} + O_P(n^{-1/2-\delta}).$$
  - If $k \sim n^{1-2\delta}$ or $k \gg n^{1-2\delta}$,
$$\tilde{\mu}'_{zx} W \tilde{\mu}_{z\varepsilon} = \frac{k}{n}\mathbb{C}_{1k} + O_P(n^{-1}\sqrt{k}).$$

We summarize parts I. and II. by claiming that:

- If $\delta \geq 1/2$,
$$\tilde{\mu}'_{zx} W \tilde{\mu}_{zx} = \frac{k}{n}\mathbb{V}_{1k} + O_P\left(\frac{\sqrt{k}}{n}\right), \qquad \tilde{\mu}'_{zx} W \tilde{\mu}_{z\varepsilon} = \frac{k}{n}\mathbb{C}_{1k} + O_P\left(\frac{\sqrt{k}}{n}\right).$$

- If $0 \leq \delta < 1/2$,
  - If $k \ll n^{1/2-\delta}$ or $k \sim n^{1/2-\delta}$,
$$\tilde{\mu}'_{zx} W \tilde{\mu}_{zx} = n^{-2\delta} a'_k W a_k + O_P\left(\frac{1}{n^{\delta+1/2}}\right), \qquad \tilde{\mu}'_{zx} W \tilde{\mu}_{z\varepsilon} = O_P(n^{-1/2-\delta}).$$
  - If $n^{1/2-\delta} \ll k \ll n^{1-2\delta}$,
$$\tilde{\mu}'_{zx} W \tilde{\mu}_{zx} = n^{-2\delta} a'_k W a_k + O_P\left(\frac{k}{n}\right), \qquad \tilde{\mu}'_{zx} W \tilde{\mu}_{z\varepsilon} = \frac{k}{n}\mathbb{C}_{1k} + O_P(n^{-1/2-\delta}).$$
  - If $k \sim n^{1-2\delta}$,
$$\tilde{\mu}'_{zx} W \tilde{\mu}_{zx} = n^{-2\delta}(\mathbb{V}_{1k} + a'_k W a_k) + O_P\left(\frac{1}{n^{\delta+1/2}}\right), \qquad \tilde{\mu}'_{zx} W \tilde{\mu}_{z\varepsilon} = \frac{k}{n}\mathbb{C}_{1k} + O_P(n^{-1}\sqrt{k}).$$
  - If $k \gg n^{1-2\delta}$,
$$\tilde{\mu}'_{zx} W \tilde{\mu}_{zx} = \frac{k}{n}\mathbb{V}_{1k} + O_P\left(\frac{\sqrt{k}}{n} \vee \frac{1}{n^{2\delta}}\right), \qquad \tilde{\mu}'_{zx} W \tilde{\mu}_{z\varepsilon} = \frac{k}{n}\mathbb{C}_{1k} + O_P(n^{-1}\sqrt{k}).$$

III. We now assess the effect of replacing W by $\hat{W}$. We use the fact that $\tilde{\mu}_{zx} = O_P(n^{-\delta} \vee \sqrt{k/n})$, $\tilde{\mu}_{z\varepsilon} = O_P(\sqrt{k/n})$, and $\hat{W} - W = o_P(k^{-1/2})$ to claim that

$$\left\| \tilde{\mu}_{zx}'(\hat{W} - W)\tilde{\mu}_{zx} \right\|_2 = O_P\left(\frac{1}{n^{2\delta}\sqrt{k}} \vee \frac{\sqrt{k}}{n}\right), \quad \text{and} \quad \left\| \tilde{\mu}_{zx}'(\hat{W} - W)\tilde{\mu}_{z\varepsilon} \right\|_2 = O_P\left(\frac{1}{n^{2\delta+1/2}\sqrt{k}} \vee \frac{\sqrt{k}}{n^{3/2}}\right).$$

As a result, we have:

- If $\delta \geq 1/2$,

$$\tilde{\mu}_{zx}'\hat{W}\tilde{\mu}_{zx} = \frac{k}{n}\mathbb{V}_{1k} + O_P\left(\frac{\sqrt{k}}{n}\right), \qquad \tilde{\mu}_{zx}'\hat{W}\tilde{\mu}_{z\varepsilon} = \frac{k}{n}\mathbb{C}_{1k} + O_P\left(\frac{\sqrt{k}}{n}\right).$$

- If $0 \leq \delta < 1/2$,
  - If $k \ll n^{1/2-\delta}$ or $k \sim n^{1/2-\delta}$,

$$\tilde{\mu}_{zx}'\hat{W}\tilde{\mu}_{zx} = n^{-2\delta}a_k'Wa_k + O_P\left(\frac{1}{n^{2\delta}\sqrt{k}}\right), \qquad \tilde{\mu}_{zx}'\hat{W}\tilde{\mu}_{z\varepsilon} = O_P(n^{-1/2-\delta}). \qquad \text{(OA.14)}$$

  - If $n^{1/2-\delta} \ll k \ll n^{1-2\delta}$,

$$\tilde{\mu}_{zx}'\hat{W}\tilde{\mu}_{zx} = n^{-2\delta}a_k'Wa_k + O_P\left(\frac{1}{n^{2\delta}\sqrt{k}} \vee \frac{k}{n}\right), \qquad \tilde{\mu}_{zx}'\hat{W}\tilde{\mu}_{z\varepsilon} = \frac{k}{n}\mathbb{C}_{1k} + O_P(n^{-1/2-\delta}).$$

  - If $k \sim n^{1-2\delta}$,

$$\tilde{\mu}_{zx}'\hat{W}\tilde{\mu}_{zx} = n^{-2\delta}(\mathbb{V}_{1k} + a_k'Wa_k) + O_P\left(\frac{1}{n^{\delta+1/2}}\right), \qquad \tilde{\mu}_{zx}'\hat{W}\tilde{\mu}_{z\varepsilon} = \frac{k}{n}\mathbb{C}_{1k} + O_P(n^{-1}\sqrt{k}).$$

  - If $k \gg n^{1-2\delta}$,

$$\tilde{\mu}_{zx}'\hat{W}\tilde{\mu}_{zx} = \frac{k}{n}\mathbb{V}_{1k} + O_P\left(\frac{\sqrt{k}}{n} \vee \frac{1}{n^{2\delta}}\right), \qquad \tilde{\mu}_{zx}'\hat{W}\tilde{\mu}_{z\varepsilon} = \frac{k}{n}\mathbb{C}_{1k} + O_P(n^{-1}\sqrt{k}).$$

IV. In this part, we derive the order of magnitude of the inverse of $\tilde{\mu}_{zx}'\hat{W}\tilde{\mu}_{zx}$ and conclude the proof. We obtain, recalling that $\tilde{\theta} - \theta_0 = (\tilde{\mu}_{zx}'\hat{W}\tilde{\mu}_{zx})^{-1}\tilde{\mu}_{zx}'\hat{W}\tilde{\mu}_{z\varepsilon}$:

- If $\delta \geq 1/2$,

$$(\tilde{\mu}_{zx}'\hat{W}\tilde{\mu}_{zx})^{-1} = \frac{n}{k}\mathbb{V}_{1k}^{-1} + O_P\left(\frac{n}{k\sqrt{k}}\right), \qquad \tilde{\theta} - \theta_0 = \mathbb{V}_{1k}^{-1}\mathbb{C}_{1k} + O_P\left(\frac{1}{\sqrt{k}}\right).$$

- If $0 \leq \delta < 1/2$,

– If $\quad k \ll n^{1/2-\delta}$ or $k \sim n^{1/2-\delta}$,

$$(\tilde{\mu}_{zx}'\hat{W}\tilde{\mu}_{zx})^{-1} = n^{2\delta}(a_k'Wa_k)^{-1} + O_P\left(\frac{n^{2\delta}}{\sqrt{k}}\right), \qquad \tilde{\theta} - \theta_0 = O_P(n^{-1/2+\delta}). \qquad \text{(OA.15)}$$

– If $\quad n^{1/2-\delta} \ll k \ll n^{1-2\delta}$,

$$(\tilde{\mu}_{zx}'\hat{W}\tilde{\mu}_{zx})^{-1} = n^{2\delta}(a_k'Wa_k)^{-1} + O_P\left(\frac{n^{2\delta}}{\sqrt{k}} \vee \frac{kn^{4\delta}}{n}\right),$$

$$\tilde{\theta} - \theta_0 = kn^{2\delta-1}(a_k'Wa_k)^{-1}\mathbb{C}_{1k} + O_P\left(k^{1/2}n^{2\delta-1} \vee k^2 n^{4\delta-2}\right)$$

– If $\quad k \sim n^{1-2\delta}$,

$$(\tilde{\mu}_{zx}'\hat{W}\tilde{\mu}_{zx})^{-1} = n^{2\delta}(\mathbb{V}_{1k} + a_k'Wa_k)^{-1} + O_P\left(n^{3\delta-1/2}\right),$$

$$\tilde{\theta} - \theta_0 = (\mathbb{V}_{1k} + a_k'Wa_k)^{-1}\mathbb{C}_{1k} + O_P(n^{-1/2+\delta}).$$

– If $\quad k \gg n^{1-2\delta}$,

$$(\tilde{\mu}_{zx}'\hat{W}\tilde{\mu}_{zx})^{-1} = \frac{n}{k}\mathbb{V}_{1k}^{-1} + O_P\left(\frac{n}{k\sqrt{k}} \vee \frac{n^{2-2\delta}}{k^2}\right), \qquad \tilde{\theta} - \theta_0 = \mathbb{V}_{1k}^{-1}\mathbb{C}_{1k} + O_P\left(\frac{1}{\sqrt{k}} \vee \frac{n^{1-2\delta}}{k}\right).$$

This concludes the proof. $\square$

OA.2 Additional results for the GMM estimator under H_1

This section presents the auxiliary Lemma OA.1 and Propositions OA.2 and OA.3. The two Propositions are related to the limiting behavior of the first-step GMM estimator and its optimal weighting matrix under H_1. The statements of the results are followed by their respective proofs.

Lemma OA.1 *Let $(U_i, z_i) \in \mathbb{R} \times \mathbb{R}^m$ be a sequence of i.i.d. random variables. Let $Z_i := g^{(k)}(z_i) \in \mathbb{R}^k$ with mean μ_z and let*

$$\hat{V}_u = \frac{1}{n}\sum_{i=1}^{n} U_i(Z_i - \mu_z)(Z_i - \mu_z)', \quad \textit{and} \quad V_u = E(U_i(Z_i - \mu_z)(Z_i - \mu_z)').$$

If $k^{-2}\sum_{l,m=1}^{k} Var\left(U_i(Z_{il} - \mu_{zl})(Z_{im} - \mu_{zm})\right) \leq \Delta < \infty$, then

$$\|\hat{V}_u - V_u\|_2 = O_P(k/\sqrt{n}).$$

Proposition OA.2 *Suppose Assumptions 1, 2, 3, and 4(a) hold, and $k \to \infty$ with $k \sim a(\log n)^b$, for some $a, b > 0$. Then, under H_1, we have:*

(a) *For* $0 \le \delta < 1/2$, $\tilde{\theta} = \theta_0 + n^{\delta}(a_k' W a_k)^{-1} a_k' W c_z + O_P\left(\frac{n^{\delta}}{\sqrt{k}}\right)$.

(b) *For* $\delta \ge 1/2$, $\tilde{\theta} = \theta_0 + O_P\left(\frac{\sqrt{n}}{k}\right)$.

If, in addition, $Var\left((x_{ih} - \mu_{xh})(Z_i - \mu_z)\right)$ has its smallest eigenvalue uniformly bounded away from 0 for at least one $h \in \{1, \dots, p\}$, then the $O_P(\sqrt{n}/k)$ remainder has a sharp order of magnitude.

Proposition OA.3 *Let $\tilde{e} = \tilde{\theta} - \theta_0$. Suppose Assumptions 1, 2, 3, 4(a), and 5 hold, and $k \to \infty$ with $k \sim a(\log n)^b$, for some $a, b > 0$. Then, under H_1, we have:*

(a) *For* $0 < \delta < 1/2$, $\tilde{V} = V_3(\tilde{e}) + O_P(n^{\delta})$, *and* $\tilde{V}^{-1} = V_3(\tilde{e})^{-1} + O_P(n^{-3\delta})$.

(b) *For* $\delta = 0$, $\tilde{V} = V_{3,0} + O_P(1/\sqrt{k})$, *and* $\tilde{V}^{-1} = V_{3,0}^{-1} + O_P(1/\sqrt{k})$.

(c) *For* $\delta \ge 1/2$, $\tilde{V} = V_3(\tilde{e}) + O_P\left(\sqrt{n}/k\right)$, *and* $\tilde{V}^{-1} = V_3(\tilde{e})^{-1} + O_P\left(k^3/n^{3/2}\right)$.

Proof of Lemma OA.1: We have:

$$\|\hat{V}_u - V_u\|_2 = \sqrt{\lambda_{\max}[(\hat{V}_u - V_u)^2]} \le \left(\sum_{l,m=1}^{k} [(\hat{V}_u - V_u)_{l,m}]^2\right)^{1/2} := \frac{1}{\sqrt{n}}\left(\sum_{l,m=1}^{k} a_{l,m}^2\right)^{1/2},$$

with $a_{l,m} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[U_i(Z_{il} - \mu_{zl})(Z_{im} - \mu_{zm}) - E(U_i(Z_{il} - \mu_{zl})(Z_{im} - \mu_{zm}))\right]$.
Note that: $E(a_{l,m}^2) = Var(U_i(Z_{il} - \mu_{zl})(Z_{im} - \mu_{zm}))$. Hence,

$$E\left(\sum_{l,m=1}^{k} a_{l,m}^2\right) = \sum_{l,m=1}^{k} Var\left(U_i(Z_{il} - \mu_{zl})(Z_{im} - \mu_{zm})\right) = O(k^2).$$

Thus, $\sum_{l,m=1}^{k} a_{l,m}^2 = O_P(k^2)$ and the conclusion follows. □

Proof of Proposition OA.2: Recall $\tilde{\theta} = \theta_0 + (\tilde{\mu}_{zx}' \hat{W} \tilde{\mu}_{zx})^{-1}(\tilde{\mu}_{zx}' \hat{W} \tilde{\mu}_{z\varepsilon})$. From part IV of the proof of Theorem 2.1, we have:

$$\left(\tilde{\mu}_{zx}' \hat{W} \tilde{\mu}_{zx}\right)^{-1} = n^{2\delta}(a_k' W a_k)^{-1} + O_P(n^{2\delta} k^{-1/2}), \quad \text{if} \quad \delta \in [0, 1/2[; \quad \text{and} \qquad \text{(OA.16)}$$

$$\left(\tilde{\mu}_{zx}' \hat{W} \tilde{\mu}_{zx}\right)^{-1} = \frac{n}{k}\mathbb{V}_{1k}^{-1} + O_P(nk^{-3/2}), \quad \text{if} \quad \delta \ge 1/2. \qquad \text{(OA.17)}$$

I. Consider $\tilde{\mu}_{zx} W \tilde{\mu}_{z\varepsilon}$. Recall that

$$\tilde{\mu}_{zx} = \bar{\mu}_{zx} - (\bar{Z} - \mu_z)(\bar{x} - \mu_x)' = \bar{\mu}_{zx} + O_P(\sqrt{k}/n) = n^{-\delta} a_k + O_P(\sqrt{k/n}).$$

Also, under our assumptions, the conclusion of Lemma A.1(c) of Dovonon and Gospodinov (2023) holds and we have

$$\tilde{\mu}_{z\varepsilon} = \bar{\mu}_{z\varepsilon} + O_P(n^{-1}\sqrt{k}), \quad \bar{\mu}_{z\varepsilon} = O_P(\|c_z\|_2); \qquad \bar{\mu}_{z\varepsilon} = \frac{1}{n}\sum_{i=1}^{n}\alpha_i + c_z; \quad \alpha_i = (Z_i - \mu_z)\varepsilon_i - c_z. \tag{OA.18}$$

We can then claim that

$$\tilde{\mu}_{zx}' W \tilde{\mu}_{z\varepsilon} = \bar{\mu}_{zx} W \bar{\mu}_{z\varepsilon} + O_P\left(\frac{\|c_z\|_2\sqrt{k}}{n}\right). \tag{OA.19}$$

Consider $\bar{\mu}_{zx}' W \bar{\mu}_{z\varepsilon}$. Let $\tilde{x}_i = x_i - \mu_x$ and pick the h-th component of $\bar{\mu}_{zx}' W \bar{\mu}_{z\varepsilon}$ given by:

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_{ih}(Z_i-\mu_z)'W\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i(Z_i-\mu_z) = \frac{1}{n^2}\sum_{i=1}^{n}\tilde{x}_{ih}(Z_i-\mu_z)'W\sum_{i=1}^{n}\alpha_i + \frac{1}{n}\sum_{i=1}^{n}\tilde{x}_{ih}(Z_i-\mu_z)'Wc_z$$
$$:= (A1) + (A2).$$

I.1. Consider $(A1)$. We have:

$$(A1) = \frac{1}{n^2}\sum_{i=1}^{n}\tilde{x}_{ih}(Z_i-\mu_z)'W\alpha_i + \frac{1}{n^2}\sum_{i\neq j}\tilde{x}_{ih}(Z_i-\mu_z)'W\alpha_j := (A1.1) + (A1.2).$$

Consider $(A1.2)$. We have: $E(A1.2) = 0$. To evaluate the order of magnitude of $(A1.2)$, we derive as previously $E(A1.2)^2$. We have:

$$\begin{aligned} E(A1.2)^2 &= \frac{1}{n^4}\sum_{i\neq j, l\neq m} E\left(\tilde{x}_{ih}(Z_i-\mu_z)'W\alpha_j \cdot \tilde{x}_{lh}(Z_l-\mu_z)'W\alpha_m\right) \\ &= \frac{1}{n^4}\sum_{i\neq j}E(\tilde{x}_{ih}(Z_i-\mu_z)'W\alpha_j)^2 + \frac{1}{n^4}\sum_{i\neq j}E(\tilde{x}_{ih}(Z_i-\mu_z)'W\alpha_j\cdot\tilde{x}_{jh}(Z_j-\mu_z)'W\alpha_i) \\ &\quad + \frac{1}{n^4}\sum_{i\neq j, l\neq j, i\neq l}E(\tilde{x}_{ih}(Z_i-\mu_z)'W\alpha_j\cdot\tilde{x}_{lh}(Z_l-\mu_z)'W\alpha_j). \end{aligned}$$

We have:

$$\begin{aligned} E(\tilde{x}_{ih}(Z_i-\mu_z)'W\alpha_j)^2 &= E\left(\alpha_j' W \cdot x_{ih}^2(Z_i-\mu_z)(Z_i-\mu_z)'\cdot W\alpha_j\right) \\ &\leq C\cdot E(\alpha_j'\alpha_j) \leq C\cdot E(\varepsilon_j^2(Z_j-\mu_z)'(Z_j-\mu_z)) = O(k), \end{aligned}$$

for some constant $C > 0$.

$$\begin{aligned} E(\tilde{x}_{ih}(Z_i-\mu_z)'W\alpha_j\cdot\tilde{x}_{jh}(Z_j-\mu_z)'W\alpha_i) &= Etrace(W^{1/2}\tilde{x}_{jh}\alpha_j(Z_j-\mu_z)'W\tilde{x}_{ih}\alpha_i(Z_i-\mu_z)'W^{1/2}) \\ &= trace(W^{1/2}LWLW^{1/2}) \leq C\cdot trace(L^2) \leq C\cdot k\cdot\|L\|_2^2, \end{aligned}$$

for some $C > 0$ and with $L := E(\tilde{x}_{jh}\alpha_j(Z_j - \mu_z)')$. We have:

$$L = E(\varepsilon_j \tilde{x}_{jh}(Z_j - \mu_z)(Z_j - \mu_z)') - c_z E((Z_j - \mu_z)'\tilde{x}_{ih}) = O(1) + O(\|c_z\|_2)O(n^{-\delta}) = O(1),$$

where we use the fact that $\|c_z\|_2 = O(1)$. We can claim that:

$$E(\tilde{x}_{ih}(Z_i - \mu_z)'W\alpha_j \cdot \tilde{x}_{jh}(Z_j - \mu_z)'W\alpha_i) = O(k).$$

For i, j, l pairwise different, we have:

$$\begin{aligned} E(\tilde{x}_{ih}(Z_i - \mu_z)'W\alpha_j \cdot \tilde{x}_{lh}(Z_l - \mu_z)'W\alpha_j) &= E(\tilde{x}_{ih}(Z_i - \mu_z)'W\alpha_j\alpha_j'W\tilde{x}_{lh}(Z_l - \mu_z) \\ &\leq CE(\tilde{x}_{ih}(Z_i - \mu_z)') \cdot E(\tilde{x}_{ih}(Z_i - \mu_z)) = O(n^{-2\delta}), \end{aligned}$$

for some $C > 0$, where we use the fact that $E(\alpha_j\alpha_j') \leq E(\varepsilon_i^2(Z_i - \mu_z)(Z_i - \mu_z)') = O(1)$.

It follows that $E(A1.2)^2 = O(k/n^2) + O(n^{-2\delta-1})$. This shows that

$$(A1.2) = O_P\left(\frac{\sqrt{k}}{n} \vee \frac{1}{n^{\delta+1/2}}\right).$$

Hence,

$$\text{If} \quad 0 \leq \delta < 1/2, \quad (A1.2) = O_P\left(\frac{1}{n^{\delta+1/2}}\right); \quad \text{and if} \quad \delta \geq 1/2, \quad (A1.2) = O_P\left(\frac{\sqrt{k}}{n}\right). \quad \text{(OA.20)}$$

Consider $(A1.1)$. We have $En(A1.1) = E(\tilde{x}_{ih}(Z_i - \mu_z)'W\alpha_i)$.

$$\begin{aligned} E[n((A1.1) - En(A1.1)]^2 &= \frac{1}{n^2}\sum_{i=1}^{n} E\left(\tilde{x}_{ih}(Z_i - \mu_z)'W\alpha_i - E(A1.1)\right)^2 \\ &= \frac{1}{n}E\left(\tilde{x}_{ih}(Z_i - \mu_z)'W\alpha_i - E(A1.1)\right)^2 \leq \frac{1}{n}E(\tilde{x}_{ih}(Z_i - \mu_z)'W\alpha_i)^2 \\ &= \frac{1}{n}E\left(\tilde{x}_{ih}(Z_i - \mu_z)'W((Z_i - \mu_z)\varepsilon_i - c_z)\right)^2 \\ &\leq \frac{2}{n}E\left(\tilde{x}_{ih}(Z_i - \mu_z)'W(Z_i - \mu_z)\varepsilon_i\right)^2 + \frac{2}{n}E\left(\tilde{x}_{ih}(Z_i - \mu_z)'Wc_z\right)^2. \end{aligned}$$

It is not hard to see that

$$E\left(\tilde{x}_{ih}^2\varepsilon_i^2[(Z_i - \mu_z)'W(Z_i - \mu_z)]^2\right) = O(k^2)$$

and

$$E\left(\tilde{x}_{ih}(Z_i - \mu_z)'Wc_z\right)^2 \leq E(\tilde{x}_{ih}^2(Z_i - \mu_z)'W(Z_i - \mu_z)) \cdot c_z'Wc_z = O(k).$$

Hence,

$$n(A1.1) = E(\tilde{x}_{ih}(Z_i - \mu_z)'W\alpha_i) + O_P(k/\sqrt{n}) = k[k^{-1}E(\tilde{x}_{ih}(Z_i - \mu_z)'W\alpha_i)] + O_P(k/\sqrt{n}).$$

Note that $E(\tilde{x}_{ih}(Z_i-\mu_z)'W\alpha_i) = E(\tilde{x}_{ih}\varepsilon_i(Z_i-\mu_z)'W(Z_i-\mu_z)) + O(n^{-\delta})$ and it follows that,

$$n(A1.1) = k[k^{-1}E(\tilde{x}_{ih}\varepsilon_i(Z_i-\mu_z)'W(Z_i-\mu_z))] + O(n^{-\delta}\vee k/\sqrt{n}).$$

We conclude that:

$$(A1.1) = \frac{k}{n}[k^{-1}E(\tilde{x}_{ih}\varepsilon_i(Z_i-\mu_z)'W(Z_i-\mu_z))] + O(n^{-1-\delta}\vee k/n^{3/2}) = O\left(\frac{k}{n}\right). \quad \text{(OA.21)}$$

From (OA.20) and (OA.21), we have:

$$\text{If} \quad 0 \le \delta < 1/2, \quad (A1) = O_P\left(\frac{1}{n^{\delta+1/2}}\right); \quad \text{and if} \quad \delta \ge 1/2, \quad (A1) = O_P\left(\frac{k}{n}\right). \quad \text{(OA.22)}$$

It follows using (OA.19) that:

$$\text{If} \quad 0 \le \delta < 1/2, \quad \tilde{\mu}_{zx}'W\tilde{\mu}_{z\varepsilon} = (A2) + O_P\left(\frac{1}{n^{\delta+1/2}}\right); \quad \text{and}$$

$$\text{if} \quad \delta \ge 1/2, \quad \tilde{\mu}_{zx}'W\tilde{\mu}_{z\varepsilon} = (A2) + O_P\left(\frac{k}{n}\right), \quad \text{(OA.23)}$$

with (we keep the same notation for the vector) $(A2) := \frac{1}{n}\sum_{i=1}^n (x_i-\mu_x)\cdot(Z_i-\mu_z)'Wc_z$.

I.2. The effect of $\tilde{W}$. We have:

$$\tilde{\mu}_{zx}'\hat{W}\tilde{\mu}_{z\varepsilon} = \tilde{\mu}_{zx}'W\tilde{\mu}_{z\varepsilon} + \tilde{\mu}_{zx}'(\hat{W}-W)\tilde{\mu}_{z\varepsilon}.$$

But

$$\|\tilde{\mu}_{zx}'(\hat{W}-W)\tilde{\mu}_{z\varepsilon}\|_2 \le \|\tilde{\mu}_{zx}\|_2\|\hat{W}-W\|_2\|\tilde{\mu}_{z\varepsilon}\|_2 = O_P(n^{-\delta}\vee\sqrt{k/n})o_P(1/\sqrt{k})O_P(\|c_z\|_2).$$

Then, from (OA.23), we have:

$$\text{If} \quad 0 \le \delta < 1/2, \quad \tilde{\mu}_{zx}'\hat{W}\tilde{\mu}_{z\varepsilon} = (A2) + o_P\left(\frac{1}{n^{\delta}\sqrt{k}}\right); \quad \text{and}$$

$$\text{if} \quad \delta \ge 1/2, \quad \tilde{\mu}_{zx}'\hat{W}\tilde{\mu}_{z\varepsilon} = (A2) + o_P\left(\frac{1}{\sqrt{n}}\right). \quad \text{(OA.24)}$$

II. Sharp order of magnitude of $(A2)$. We have

$$E(A2) = E(\tilde{x}_{ih}(Z_i-\mu_z)'Wc_z) = n^{-\delta}a_{kh}'Wc_z,$$

where a_{kh} is the h-th component of a_k as defined in Assumption 1. To check the order of magnitude of $(A2)$, we derive:

$$E((A2) - E(A2))^2 =$$
$$E\left(\frac{1}{n}\sum_{i=1}^{n}(\tilde{x}_{ih}(Z_i - \mu_z)'Wc_z - n^{-\delta}a'_{kh}Wc_z)\right)^2 = \frac{1}{n}E(\tilde{x}_{ih}(Z_i - \mu_z)'Wc_z - n^{-\delta}a'_{kh}Wc_z)^2$$
$$= \frac{1}{n}c'_zWE\left((\tilde{x}_{ih}(Z_i - \mu_z) - n^{-\delta}a_{kh}) \cdot (\tilde{x}_{ih}(Z_i - \mu_z) - n^{-\delta}a_{kh})'\right)Wc_z$$
$$= \frac{1}{n}c'_zWVar(\tilde{x}_{ih}(Z_i - \mu_z))Wc_z.$$

Under the condition that $Var(\tilde{x}_{ih}(Z_i - \mu_z))$ has its smallest eigenvalue uniformly bounded away from 0, we can claim that:

$$(C_1/n)\|c_z\|_2^2 \leq E((A2) - E(A2))^2 \leq (C_2/n)\|c_z\|_2^2,$$

for some constants $C_1, C_2 > 0$. Thus,

$$\text{if} \quad 0 \leq \delta < 1/2, \quad (A2) = n^{-\delta}a'_{kh}Wc_z + O_P\left(\frac{1}{\sqrt{n}}\right); \quad \text{and}$$
$$\text{if} \quad \delta \geq 1/2, \quad (A2) = O_P\left(\frac{1}{\sqrt{n}}\right), \quad \text{(OA.25)}$$

where the $O_P(1/\sqrt{n})$ bound is sharp in each case.

III. Using (OA.16), (OA.17), (OA.24) and (OA.25), some straightforward derivations yield:

If $0 \leq \delta < 1/2$,

$$\tilde{\theta} = \theta_0 + n^{\delta}(a'_kWa_k)^{-1}a'_kWc_z + O_P\left(\frac{n^{\delta}}{\sqrt{k}}\right); \quad \text{(OA.26)}$$

and if $\delta \geq 1/2$,

$$\tilde{\theta} = \theta_0 + O_P\left(\frac{\sqrt{n}}{k}\right), \quad \text{(OA.27)}$$

where the $O_P(\sqrt{n}/k)$ term in the case $\delta \geq 1/2$ is sharp. $\square$

Proof of Proposition OA.3: We have: $\tilde{V} = \frac{1}{n}\sum_{i=1}^{n}\tilde{\varepsilon}_i^2(Z_i - \bar{Z})(Z_i - \bar{Z})'$. We observe that

$$\tilde{\varepsilon}_i = -(x_i - \mu_x)'(\tilde{\theta} - \theta_0) + (\bar{x} - \mu_x)'(\tilde{\theta} - \theta_0) + \varepsilon_i - \bar{\varepsilon}.$$

From Proposition OA.2: For $0 < \delta < 1/2$, $\tilde{e} := \tilde{\theta} - \theta_0 = O_P(n^{\delta})$;

for $\delta = 0$, $\tilde{e} := \tilde{\theta} - \theta_0 - b_{k,0} = O_P(1/\sqrt{k})$, with $b_{k,0} = (a'_kWa_k)^{-1}a'_kWc_z$; and

for $\quad \delta \geq 1/2, \quad \tilde{e} := \tilde{\theta} - \theta_0 = O_P(\sqrt{n}/k)$. Using this, we find that:
If $0 < \delta < 1/2$,

$$\tilde{V} = \frac{1}{n}\sum_{i=1}^{n} \tilde{\varepsilon}_i^2 (Z_i - \mu_z)(Z_i - \mu_z)' + O_P\left(kn^{2\delta - 1/2}\right).$$

If $\delta = 0$,

$$\tilde{V} = \frac{1}{n}\sum_{i=1}^{n} \tilde{\varepsilon}_i^2 (Z_i - \mu_z)(Z_i - \mu_z)' + O_P\left(kn^{-1/2}\right).$$

Finally, if $\delta \geq 1/2$

$$\tilde{V} = \frac{1}{n}\sum_{i=1}^{n} \tilde{\varepsilon}_i^2 (Z_i - \mu_z)(Z_i - \mu_z)' + O_P\left(k^{-1}\sqrt{n}\right).$$

(a) Consider the case: $0 < \delta < 1/2$. Straightforward calculations yields:

$$\frac{1}{n}\sum_{i=1}^{n} \tilde{\varepsilon}_i^2 (Z_i - \mu_z)(Z_i - \mu_z)' = V_3(\tilde{e}) + O_P(n^{\delta})$$

so that

$$\tilde{V} = V_3(\tilde{e}) + O_P(n^{\delta}).$$

Assumptions 5 and 4(a) ensure that

$$\underline{\lambda} \leq \inf_{u \in \mathbb{R}^p : \|u\|_2 = 1} \lambda_{\min}(V_3(u)) \leq \sup_{u \in \mathbb{R}^p : \|u\|_2 = 1} \lambda_{\max}(V_3(u)) \leq C\bar{\lambda},$$

for some positive constant C. Therefore, we can claim, since $V_3(u)$ is an homogeneous function of degree 2, that

$$\underline{\lambda} \cdot \|\tilde{e}\|_2^2 \cdot I_k \leq V_3(\tilde{e}) \leq C\bar{\lambda} \cdot \|\tilde{e}\|_2^2 \cdot I_k,$$

where the inequality is in terms of matrices: ($A \leq B$ meaning that $A - B$ is positive semidefinite.)

Since the leading term in the expansion of $\|\tilde{e}\|^2$ is $n^{2\delta} c_z' W a_k (a_k' W a_k)^{-2} a_k' W c_z$, we use the fact that: $\liminf_k \|a_k' W c_z\|_2 > 0$, $\liminf_k \lambda_{\min}(a_k' a_k) > 0$ and the fact that W has bounded eigenvalues to claim that

$$\|\tilde{e}\|_2^{-2} = O_P(n^{-2\delta}).$$

It follows that

$$V_3(\tilde{e})^{-1} = O_P(n^{-2\delta}).$$

Now, write $\tilde{V} = V_3(\tilde{e}) + \mathcal{E}$, with $\mathcal{E} = O_P(n^{\delta})$. We have

$$\|V_3(\tilde{e})^{-1}\mathcal{E}\|_2 = O_P(n^{-\delta}) = o_P(1).$$

Thus,

$$(I_k + V_3(\tilde{e})^{-1}\mathcal{E})^{-1} = I_k - V_3(\tilde{e})^{-1}\mathcal{E} + O_P(n^{-2\delta})$$

and

$$\tilde{V}^{-1} = V_3(\tilde{e})^{-1} - V_3(\tilde{e})^{-1}\mathcal{E}V_3(\tilde{e})^{-1} + O_P(n^{-4\delta}) = V_3(\tilde{e})^{-1} + O_P(n^{-3\delta}).$$

This completes the proof of part (a).

(b) Consider the case $\delta = 0$. Using the expression of $\tilde{\varepsilon}_i$ in which we replace $\tilde{\theta} - \theta_0$ by $\tilde{e} + b_{k,0}$, with, as we recalled $\tilde{e} = O_P(1/\sqrt{k})$ (note also that $b_{k,0} = O(1)$ which ensures along with the conditions on the relevant eigenvalues that $V_{3,0}$ is bounded), we obtain:

$$\tilde{V} = V_{3,0} + O_P(1/\sqrt{k}), \quad \text{implying that} \quad \tilde{V}^{-1} = V_{3,0}^{-1} + O_P(1/\sqrt{k}).$$

(c) Consider the case $\delta \geq 1/2$. The proof of this part follows the same lines as the proof of part (a). We obtain:

$$\tilde{V} = V_3(\tilde{e}) + O_P\left(\frac{\sqrt{n}}{k}\right).$$

The leading term of $\tilde{e}$ can be obtained from the proof of Proposition OA.2 as

$$\frac{n}{k}\mathbb{V}_{1k}^{-1}\frac{1}{n}\sum_{i=1}^{n}\left[(x_i - \mu_x)(Z_i - \mu_z)' - E((x_i - \mu_x)(Z_i - \mu_z)')\right]Wc_z.$$

By application of the central limit theorem for independent and (row-wise) identically distributed triangular arrays, we can claim that

$$n^{-1/2}\sum_{i=1}^{n}\left[(x_i - \mu_x)(Z_i - \mu_z)'Wc_z - E((x_i - \mu_x)(Z_i - \mu_z)'Wc_z)\right]$$

is asymptotically normally distributed and, therefore, does not have an atom mass at 0 as n grows if the asymptotic variance is nondegenerate. Actually, we can make the claim of no atom of probability at 0 if at least one component, say h of the quantity above is asymptotically normal with nondegenerate variance. This is the case for any component h such that the smallest eigenvalue of $Var\left((x_{ih} - \mu_{xh})(Z_i - \mu_z)\right)$ is uniformly bounded away from 0. This shows that $\tilde{e}$ properly scaled by its order of magnitude does not have any atom at 0. It follows that, $\|\tilde{e}\|_2^{-2} = O_P(k^2/n)$ which allows to claim the stated result. □

OA.3 Lower bound on Δ_k in Theorem 4.2

The aim of this appendix is to show that the sequence Δ_k, appearing in Theorem 4.2, is uniformly bounded away from 0 so that consistency of the proposed test is ensured. For this purpose, Δ_k is considered the image of a specific point in $\mathbb{R}^p$ of a function that we study in Lemmas OA.4 and OA.5 below. The connection to Theorem 4.2 is established in Remark OA.1 below.

Let $\Sigma_k(u)$ be a sequence of symmetric positive definite (k,k)-matrices such that for any $k \geq 2$,

$$\Sigma_{k+1}(u) = \begin{pmatrix} \Sigma_k(u) & v \\ v' & \sigma^2 \end{pmatrix},$$

where $v \in \mathbb{R}^k$ and $\sigma^2 > 0$ are also functions of $u \in \mathbb{R}^p$, but we omit the dependence for simplicity. Let a_k and c_k be two sequences of (k,p)-matrices and $(k,1)$-vectors, respectively, such that a_k is full rank for k large enough and

$$a_{k+1} = \begin{pmatrix} a_k \\ \alpha'_{k+1} \end{pmatrix}, \quad c_{k+1} = \begin{pmatrix} c_k \\ \beta'_{k+1} \end{pmatrix}$$

for some sequences $\alpha_k \in \mathbb{R}^k$ and $\beta_k \in \mathbb{R}$. Define

$$P_k := \Sigma_k(u)^{-1/2} a_k \left(a'_k \Sigma_k(u)^{-1} a_k\right)^{-1} a'_k \Sigma_k(u)^{-1/2}, \quad \text{and} \quad \Delta_k(u) := c'_k \Sigma_k(u)^{-1/2} \left(I_k - P_\delta\right) \Sigma_k(u)^{-1/2} c_k.$$

Lemma OA.4 *$u \mapsto \Delta_k(u)$ is a non-decreasing sequence of functions, i.e., for any $k \geq 2$,*

$$\Delta_k(u) \leq \Delta_{k+1}(u), \quad \forall u \in \mathbb{R}^p.$$

Lemma OA.5 *Assume that the sequence of real-valued functions $u \mapsto \Delta_k(u)$ are all defined on a compact subset $\mathcal{C}$ of $\mathbb{R}^p$. If there exists k_0 such that $u \mapsto \Delta_{k_0}(u)$ is continuous and a_{k_0} does not lie in the column span of a_{k_0}, then, for any $u \in \mathcal{C}$ and any $k \geq k_0$,*

$$\Delta_k(u) \geq \Delta_{k_0}(u) \geq \underline{\nu},$$

where $\underline{\nu} > 0$ is an absolute constant.

Remark OA.1 *Lemma OA.5 applies to Δ_k introduced in Theorem 4.2. Indeed, note that, for $\delta = 0$, Δ_k is obtained through $\Sigma_k(u) = E((\varepsilon_i - u'(x_i - \mu_x))^2 (Z_i - \mu_z)(Z_i - \mu_z)')$, with $u = b_{0,k} = (a'_k W a_k)^{-1}(a'_k W c_z)$. Also, note that:*

$$\|b_{0,k}\|_2 \leq \frac{\|a_k\|_2 \lambda_{\max}(W) \|c_z\|_2}{\lambda_{\min}(W)\lambda_{\min}(a'_k a_k)} \leq M,$$

where $M > 0$ is an absolute constant. The last inequality holds under Assumptions 1(b) and 3 and the conditions $\|c_z\|_2 = O(1)$ and $\lambda_{\min}(W) > \underline{\lambda} > 0$. For the result in Lemma OA.5 to hold, we can consider u to lie in the compact: $\mathcal{C} = \{u \in \mathbb{R}^p : \|u\|_2 \leq M\}$.

In the case $0 < \delta < 1/2$, Δ_k is obtained through $\Sigma_k(u) = E((u'(x_i - \mu_x))^2(Z_i - \mu_z)(Z_i - \mu_z)')$, with $u = \tilde{u} = (\tilde{\theta} - \theta_0)/\|\tilde{\theta} - \theta_0\|_2$. We can then consider u to lie in the compact: $\mathcal{C} = \{u \in \mathbb{R}^p : \|u\|_2 = 1\}$.

In both cases, $\Sigma_k(u)$ is positive definite for all $u \in \mathcal{C}$ and continuous in u as a polynomial function. It follows that $u \mapsto \Delta_k(u)$ is continuous and the conditions of Lemma OA.5 are satisfied. We can therefore conclude in both cases that $\Delta_k \geq \underline{\nu} > 0$ for k large enough and this, so long as as there exists k_0 such that $c_z (\in \mathbb{R}^{k_0})$ does not belong to the column span of a_{k_0}.

Proof of Lemma OA.4: We have:

$$\Delta_{k+1}(u) := c'_{k+1}\Sigma_{k+1}(u)^{-1}c_{k+1} - c'_{k+1}\Sigma_{k+1}(u)^{-1}a_{k+1}\left(a'_{k+1}\Sigma_{k+1}(u)^{-1}a_{k+1}\right)^{-1}a'_{k+1}\Sigma_{k+1}(u)^{-1}c_{k+1}.$$

Using the block inverse formula, we get:

$$\Sigma_{k+1}(u)^{-1} = \begin{pmatrix} \Sigma_k(u)^{-1} + \Sigma_k(u)^{-1}vv'\Sigma_k(u)^{-1}/\mu & -\Sigma_k(u)^{-1}v/\mu \\ -v'\Sigma_k(u)^{-1}/\mu & 1/\mu \end{pmatrix}, \quad \mu = \sigma^2 - v'\Sigma_k(u)^{-1}v.$$

Straightforward derivations yield:

$$a'_{k+1}\Sigma_{k+1}(u)^{-1}a_{k+1} = a'_k\Sigma_k(u)^{-1}a_k + \frac{1}{\mu}VV', \text{ with } V = a'_k\Sigma_k(u)^{-1}v - \alpha'_{k+1}.$$

Let us define $S_k := a'_k\Sigma_k(u)^{-1}a_k$. By the Woodbury formula, we get:

$$S_{k+1}^{-1} := \left(a'_{k+1}\Sigma_{k+1}(u)^{-1}a_{k+1}\right)^{-1} = S_k^{-1} - \frac{S_k^{-1}\cdot\frac{1}{\mu}VV'\cdot S_k^{-1}}{1 + \frac{1}{\mu}V'S_k^{-1}V}. \tag{OA.28}$$

Also, we obtain:

$$a'_{k+1}\Sigma_{k+1}(u)^{-1}c_{k+1} = a'_k\Sigma_k(u)^{-1}c_k + \frac{\lambda_k}{\mu}V, \quad \lambda_k = v'\Sigma_k(u)^{-1}c_k - \beta_{k+1}, \tag{OA.29}$$

and

$$c'_{k+1}\Sigma_{k+1}(u)^{-1}c_{k+1} = c'_k\Sigma_k(u)^{-1}c_k + \frac{\lambda_k^2}{\mu}. \tag{OA.30}$$

Plugging (OA.28), (OA.29) and (OA.30) into the expression of $\Delta_{k+1}(u)$ above, we obtain:

$$\begin{aligned}\Delta_{k+1}(u) = \Delta_k(u) - \frac{2\lambda_k}{\mu}\left(c'_k\Sigma_k(u)^{-1}a_k\right)S_k^{-1}V - \frac{\lambda_k^2}{\mu^2}V'S_k^{-1}V + \frac{\lambda_k^2}{\mu} \\ + \left(c'_k\Sigma_k(u)^{-1}a_k + \frac{\lambda_k}{\mu}V'\right)\frac{S_k^{-1}\cdot\frac{1}{\mu}VV'\cdot S_k^{-1}}{1 + \frac{1}{\mu}V'S_k^{-1}V}\left(a'_k\Sigma_k(u)^{-1}c_k + \frac{\lambda_k}{\mu}V\right).\end{aligned}$$

Setting $A_1 = c_k'\Sigma_k(u)^{-1}a_k S_k^{-1}V$ and $A_2 = V'S_k^{-1}V$, we get:

$$\Delta_{k+1}(u) = \Delta_k(u) - \frac{2\lambda_k}{\mu}A_1 - \frac{\lambda_k^2}{\mu^2}A_2 + \frac{\lambda_k^2}{\mu} + \frac{\frac{1}{\mu}\left(A_1 + \frac{\lambda_k}{\mu}A_2\right)^2}{1+\frac{1}{\mu}A_2} = \Delta_k(u) + \frac{(A_1-\lambda_k)^2}{\mu + A_2} \geq \Delta_k(u). \quad \square$$

Proof of Lemma OA.5: From Lemma OA.4, $\Delta_k(u) \geq \Delta_{k_0}(u)$ for all $u \in \mathcal{C}$ and all $k \geq k_0$. Since $u \mapsto \Delta_{k_0}(u)$ is continuous on the compact set $\mathcal{C}$, there exists $u_0 \in \mathcal{C}$ such that $\Delta_{k_0}(u_0) = \min_{u\in\mathcal{C}} \Delta_{k_0}(u)$. The fact that c_{k_0} does not belong to the column span of a_{k_0} ensures that $\Delta_{k_0}(u_0) > 0$. We can take $\underline{\delta} = \Delta_{k_0}(u_0)$. $\square$

OA.4 Limiting distribution of the GMM estimator for $0 \leq \delta < 1/2$

Let $\Sigma_k := (a_k'Wa_k)^{-1}a_k'WE[\varepsilon_i^2(Z_i-\mu_z)(Z_i-\mu_z)')]Wa_k(a_k'Wa_k)^{-1}$.

Proposition OA.6 *Under H_0, if Assumptions 1, 2 and 3 hold for $0 \leq \delta < 1/2$, $k \sim a(\ln n)^b$, for $a, b > 0$ and Σ_k converges to Σ as $k \to \infty$, then*

$$n^{1/2-\delta}(\tilde{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma).$$

As it appears, the possibility of using the result in Proposition OA.6 to carry out feasible and asymptotically valid inference about θ_0 may rely on knowing δ and a useful estimate of a_k. Nevertheless, it turns out that this is not necessary. From the definition of a_k in Assumption 1, we also have

$$a_k = n^\delta E((Z_i - \mu_z)(x_i - \mu_x)') := n^\delta \ell_k.$$

Hence, we can claim that:

$$\sqrt{n}(\tilde{\theta} - \theta_0) \sim AN(0, n^{2\delta}\Sigma_k), \text{ and } n^{2\delta}\Sigma_k = (\ell_k'W\ell_k)^{-1}\ell_k'WE[\varepsilon_i^2(Z_i-\mu_z)(Z_i-\mu_z)')]W\ell_k(\ell_k'W\ell_k)^{-1}.$$

With this observation, the crucial observation of Antoine and Renault (2009) that there is no need to know δ to carry out inference about θ also holds in our context. Standard GMM inference formulas are obtained by replacing the quantities in the expression of $n^{2\delta}\Sigma_k$ above by their respective sample analogues. It is not hard to see that the standard GMM inference formulas lead to valid inference in the context of Proposition OA.6, regardless of the value of $\delta \in [0, 1/2)$.

Remark OA.2 *In contrast to the result in Antoine and Renault (2009), there is no parameter space rotation involved in the asymptotic distribution of the GMM estimator $\tilde{\theta}$. This is mainly* *due to the fact that each component of $E(x_i - \mu_x|z_i)$ is local-to-zero at the same rate $n^{-\delta}$. Such a rotation may appear in the asymptotic distribution of $\tilde{\theta}$ if each component of x_i is allowed to have a conditional mean with a specific local-to-zero rate. Although, this is not expected to affect the distribution of the specification test (see Section 2.1), this may change the distribution derived in Proposition OA.6 for the parameter estimate. The full study of the GMM estimator in this identification configuration is beyond the scope of this paper.*

Proof of Proposition OA.6: Recall that $\tilde{\theta} = \theta_0 + \left(\tilde{\mu}'_{zx}\hat{W}\tilde{\mu}_{zx}\right)^{-1}\left(\tilde{\mu}'_{zx}\hat{W}\tilde{\mu}_{z\varepsilon}\right)$. We have:

$$\tilde{\mu}_{zx} = n^{-\delta}a_k + O_P(\sqrt{k/n}) = O_P(n^{-\delta}), \quad \text{and} \quad \tilde{\mu}_{z\varepsilon} = \bar{\mu}_{z\varepsilon} + O_P(\sqrt{k}/n) = O_P(\sqrt{k/n}).$$

Using the fact that $\hat{W} - W = o_P(k^{-1/2})$ we obtain:

$$\tilde{\mu}'_{zx}\hat{W}\tilde{\mu}_{z\varepsilon} = n^{-\delta}a'_k W\bar{\mu}_{z\varepsilon} + o_P(n^{-1/2-\delta}).$$

Using (OA.15), we obtain:

$$n^{1/2-\delta}\left(\tilde{\mu}'_{zx}\hat{W}\tilde{\mu}_{zx}\right)^{-1}\left(\tilde{\mu}'_{zx}\hat{W}\tilde{\mu}_{z\varepsilon}\right) = (a'_k W a_k)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} a'_k W(Z_i - \mu_z)\varepsilon_i + o_P(1).$$

We use the fact that $\{a'_k W(Z_i - \mu_z)\varepsilon_i\}_i$ is a triangular array that is row-wise i.i.d. with zero mean and variance that is uniformly bounded to claim, by the central limit theorem that:

$$n^{1/2-\delta}(\tilde{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma),$$

with

$$\Sigma = \lim_{n\to\infty}(a'_k W a_k)^{-1}a'_k W E[\varepsilon_i^2(Z_i - \mu_z)(Z_i - \mu_z)')]Wa_k(a'_k W a_k)^{-1}. \quad \square$$

OA.5 Power with single regressor and irrelevant instrument

This section illustrates in the simplest model that, in the presence of completely irrelevant instruments, the power of the proposed test originates essentially from growing number, k, of generated instruments. More precisely, in the presence of irrelevant instruments, 2SGMM diverges but with a rate that is moderated by k. Thus the inverse of the *optimal* variance does not converge to 0 as fast as the signal part of $J_{n,k}$ diverges under H_1. This favorable trade-off is at the source of power.

Consider the linear model without intercept and only one regressor, which may be endogenous, and one instrument. The absence of intercept implies that, demeaning is not essential as we can show that the test statistic in (7) is equivalent to the same expression but with $J_{n,k}$ replaced by

$$\bar{J}_{n,k} := n\left(\bar{\mu}_{zy} - \hat{\theta}\bar{\mu}_{zx}\right)' \hat{V}^{-1}\left(\bar{\mu}_{zy} - \hat{\theta}\bar{\mu}_{zx}\right), \quad \text{with} \quad \hat{V} = n^{-1}\sum_{i=1}^{n}[y_i - \hat{\theta}x_i]^2 Z_i Z_i',$$

$\bar{\mu}_{\alpha\beta} = n^{-1}\sum_{i=1}^{n}(\alpha_i - E(\alpha_i))(\beta_i - E(\beta_i))'$, and $\hat{\theta}$ the 2SGMM estimator based on a first step estimator $\tilde{\theta}$ associated to identity weighting matrix and obtained without demeaning. Recalling that $c_z := E(Z_i\varepsilon_i)$, Straightforward calculations (also see proof of Proposition OA.2(b)) yields

$$\tilde{\theta} = \theta_0 + \tilde{e} := \theta_0 + \frac{\sqrt{n}}{k}\left(\sqrt{n}\bar{\mu}_{zx}' c_z\right) \Big/ \sqrt{\frac{n}{k}}\bar{\mu}_{zx}'\sqrt{\frac{n}{k}}\bar{\mu}_{zx} + O_P(1).$$

Let $\tilde{V}$ be as $\hat{V}$ but with $\tilde{\theta}$ replacing $\hat{\theta}$ and $V_3(1) = E(x_i^2 Z_i Z_i')$. We have (see Proposition OA.3(c)),

$$\tilde{V} = \tilde{e}^2 V_3(1) + O_P(\sqrt{n}/k), \quad \text{and} \quad \tilde{V}^{-1} = \tilde{e}^{-2}(V_3(1))^{-1} + O_P(k^3/n^{3/2}).$$

Moreover,

$$\begin{aligned}
\hat{\theta} = \theta_0 + \hat{e} &:= \theta_0 + \frac{\sqrt{n}}{k}\left(\sqrt{n}\bar{\mu}_{zx}'\tilde{V}^{-1}c_z\right) \Big/ \left(\sqrt{\frac{n}{k}}\bar{\mu}_{zx}'\tilde{V}^{-1}\sqrt{\frac{n}{k}}\bar{\mu}_{zx}\right) + O_P(1) \\
&= \theta_0 + \frac{\sqrt{n}}{k}\left(\sqrt{n}\bar{\mu}_{zx}'(V_3(1))^{-1}c_z\right) \Big/ \left(\sqrt{\frac{n}{k}}\bar{\mu}_{zx}'(V_3(1))^{-1}\sqrt{\frac{n}{k}}\bar{\mu}_{zx}\right) + O_P(k^{1/2}) \\
&:= \frac{\sqrt{n}}{k}h_n + O_P(k^{1/2}).
\end{aligned}$$

We also obtain:

$$\hat{V} = \hat{e}^2 V_3(1) + O_P(\sqrt{n}/k), \quad \text{and} \quad \hat{V}^{-1} = \hat{e}^{-2}(V_3(1))^{-1} + O_P(k^3/n^{3/2}).$$

Turning back to the signal part of $\bar{J}_n$, we obtain

$$\sqrt{n}(\bar{\mu}_{zy} - \hat{\theta}\bar{\mu}_{zx}) = \sqrt{n}(\bar{\mu}_{z\varepsilon} - \hat{e}\bar{\mu}_{zx}) = \sqrt{n}c_z + O_P(\sqrt{n/k}) = O_P(\sqrt{n}).$$

It follows that:

$$\bar{J}_n = n(c_z + O_P(1/\sqrt{k}))'\hat{e}^{-2}V_3(1)^{-1}(c_z + O_P(1/\sqrt{k})) + O_P(k^3/\sqrt{n}).$$

Noting that $\hat{e}^{-2} = (n^{1/2}k^{-1}h_n)^{-2} + O_P(k^{7/2}/n^{3/2})$ and that both h_n and h_n^{-1} are $O_P(1)$, we obtain

$$\bar{J}_n = k^2 h_n^{-2} c_z'(V_3(1))^{-1}c_z + O_P(k^{3/2}).$$

Since $\|c_z\|_2$ is bounded away from 0 (under H_1), the leading term of $\bar{J}_n$ is $k^2 h_n^{-2} c_z'(V_3(1))^{-1}c_z$, which diverges to infinite as n grows. Turning to our test statistic $S_{n,k}$, its leading term, by definition is $2^{-1/2}k^{3/2}h_n^{-2}c_z'(V_3(1))^{-1}c_z$ and this ensures that $S_{n.k}$ diverges to $+\infty$ with probability approaching 1 as n grows. This establishes that the test is consistent. By the expression of this leading term, consistency of the test is essentially due to increasing k. For k fixed, this argument would not hold.

OA.6 Additional simulation evidence

This section reports additional simulation results for other sample sizes n and different distributional assumptions. Figure OA.1 presents the power curves for the $S_{n,k}$ and J_n tests at the 5% nominal level but for $n = 5,000$ observations instead of $n = 500$ as in Figure 2 in the paper.

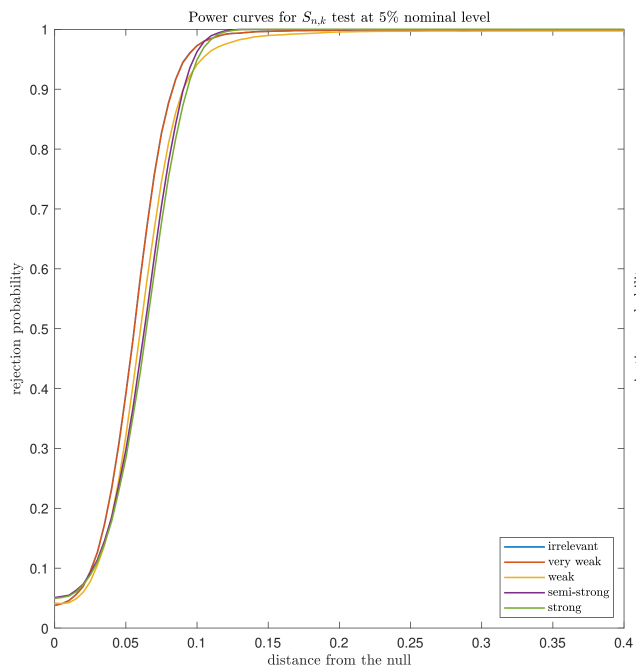

FIGURE OA.1. Empirical power curves at 5% nominal level of the $S_{n,k}$ test (left chart) and the J_n test (right chart) for various degrees of the identification signal: 'irrelevant' ($\delta = 100$), 'very weak' ($\delta = 1$), 'weak' ($\delta = 0.5$), 'semi-strong' ($\delta = 0.2$), and 'strong' ($\delta = 0$). The sample size is $n = 5,000$.

Table OA.1 is based on the same simulation design as Table 1 in the main text but $n = 100$ instead of $n = 500$ as in Table 1. While the size distortions of the test increase in the extreme tails of the distribution, the test $S_{n,k}$ exhibits only mild over-rejections at 10% nominal level. Table OA.2 is based on the same simulation design as Table 1 in the main text with $n = 500$ but $(\varepsilon_i, v_i)'$ is bivariate t-distributed with the same variance matrix Ω and 5 degrees of freedom and z_i is also drawn independently from a t-distribution with 5 degrees of freedom and I_3 as a variance matrix. This design is to assess the robustness of the proposed test to violations of Assumption 2(a).

Table OA.1. Empirical rejection rates (size and power) of the $S_{n,k}$ test with instruments that exhibit differential identification strength as a function of $(\delta_1, \delta_2, \delta_3, \delta_4)$ with $n = 100$.

	Panel A: one-sided $S_{n,k}$ test						Panel B: two-sided $S_{n,k}$ test					
$(\delta_1, \delta_2, \delta_3, \delta_4)$	size			power			size			power		
	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$(0, 0.5, 0.2, 100)$	3.3	8.2	12.8	42.7	59.1	67.6	2.3	5.9	10.6	37.0	51.4	59.2
$(100, 0.3, 0.1, 100)$	2.9	7.3	11.6	50.4	65.2	72.4	2.0	5.3	10.2	45.1	58.5	65.4
$(0, 0.2, 0.5, 0)$	3.2	8.2	12.8	55.0	73.1	81.5	2.3	5.8	10.5	48.3	64.8	73.1
$(0.8, 0.2, 0.5, 0.4)$	2.7	6.7	10.7	52.6	66.4	73.5	1.8	5.1	10.1	47.4	60.2	66.6
$(0.5, 0.4, 0.3, 0.1)$	2.6	6.7	10.6	45.1	61.2	69.8	1.8	5.1	9.9	39.5	53.8	61.4
$(0, 100, 100, 0)$	3.4	8.5	13.2	60.2	77.2	84.8	2.4	6.0	10.8	53.7	69.6	77.2
$(0.1, 0.2, 0.5, 0.5)$	2.5	6.7	10.6	30.2	45.5	54.6	1.8	5.0	9.9	25.4	38.2	45.9
$(0.6, 0.5, 0.2, 1)$	2.5	6.3	10.1	48.1	61.7	68.8	1.7	4.9	9.7	43.2	55.6	62.0

Table OA.2. Empirical rejection rates (size and power) of the $S_{n,k}$ test with instruments that exhibit differential identification strength as a function of $(\delta_1, \delta_2, \delta_3, \delta_4)$ with $n = 500$ and data drawn from a multivariate t-distribution with 5 degrees of freedom.

	Panel A: one-sided $S_{n,k}$ test						Panel B: two-sided $S_{n,k}$ test					
$(\delta_1, \delta_2, \delta_3, \delta_4)$	size			power			size			power		
	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$(0, 0.5, 0.2, 100)$	1.5	4.6	8.1	93.7	96.9	97.9	1.0	3.9	8.5	92.1	95.5	96.9
$(100, 0.3, 0.1, 100)$	1.4	4.5	7.9	89.0	93.1	94.9	0.9	3.9	8.7	87.3	91.4	93.1
$(0, 0.2, 0.5, 0)$	1.3	4.4	7.8	100	100	100	0.8	3.6	8.1	100	100	100
$(0.8, 0.2, 0.5, 0.4)$	1.2	4.0	7.0	93.2	95.9	96.9	0.8	3.7	8.9	92.1	94.8	95.9
$(0.5, 0.4, 0.3, 0.1)$	1.0	3.6	6.4	95.6	97.4	98.1	0.7	3.6	8.6	94.8	96.6	97.4
$(0, 100, 100, 0)$	1.3	4.4	7.8	100	100	100	0.8	3.6	8.1	100	100	100
$(0.1, 0.2, 0.5, 0.5)$	1.1	3.6	6.4	86.3	91.1	93.2	0.7	3.6	8.5	84.3	89.0	91.2
$(0.6, 0.5, 0.2, 1)$	1.1	3.7	6.5	89.6	93.2	94.8	0.7	3.6	8.8	88.1	91.7	93.2