Journal of Econometrics 240 (2024) 105647

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Locally robust inference for non-Gaussian linear simultaneous equations models☆

Adam Lee [a], Geert Mesters [b,*]

[a] BI Norwegian Business School, Norway
[b] Universitat Pompeu Fabra, Barcelona School of Economics and CREI, Spain

ARTICLE INFO

JEL classification:
C12
C14
C30

Keywords:
Weak identification
Semiparametric modeling
Independent component analysis
Simultaneous equations

ABSTRACT

All parameters in linear simultaneous equations models can be identified (up to permutation and sign) if the underlying structural shocks are independent and at most one of them is Gaussian. Unfortunately, existing inference methods that exploit such identifying assumptions suffer from size distortions when the true distributions of the shocks are close to Gaussian. To address this *weak non-Gaussian* problem we develop a locally robust semi-parametric inference method which is simple to implement, improves coverage and retains good power properties. The finite sample properties of the methodology are illustrated in a large simulation study and an empirical study for the returns to schooling.

1. Introduction

The linear simultaneous equations model (LSEM) is a benchmark model used to analyze general equilibrium relationships in economics. It was placed in its modern form by Haavelmo (1943, 1944), building on Frisch (1933) and Tinbergen (1939) among others. As is well known, without additional restrictions, not all parameters of the LSEM can be uniquely identified from the first and second moments of the observed data series, see Dhrymes (1994) for an in-depth discussion.

Interestingly, this identification problem vanishes (up to permutation and scale) when the underlying structural shocks are independent and at most one of them follows a Gaussian distribution (e.g. Comon, 1994). This identification approach has a long history in the statistics and signal processing literatures where it is often referred to as independent components analysis, see Hyvärinen et al. (2001) for a textbook treatment. More recently, this approach has been adopted in the econometrics literature, where interest has centered on developing methodology for conducting inference on the parameters of various LSEMs based on non-Gaussian identification (e.g. Gouriéroux et al., 2017).

Unfortunately, if in the true data generating process multiple structural shocks follow a Gaussian distribution some structural parameters may be under- or un-identified and standard inference methods that aim to exploit non-Gaussian distributions may fail to control size. Moreover, as is typical in models with points of identification failure, such behavior is also observed if the true distributions of the shocks are sufficiently close to Gaussianity. Intuitively, in such *weakly non-Gaussian* settings the available identifying information is limited relative to sampling variation leading to asymptotic coverage distortions when using standard inference methods, such as maximum likelihood and moment condition based methods.

☆ We thank numerous seminar participants for helpful comments. Mesters acknowledges support from the Spanish Ministry of Economy and Competitiveness through the Ramon y Cajal fellowship (RYC2019-028287-I) and the Spanish Ministry of Economy and Competitiveness through the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2019-000915-S) and The Netherlands Organization for Scientific Research (NWO) through the VENI research grant (016.Veni.195.036).

* Corresponding author.
E-mail address: geert.mesters@upf.edu (G. Mesters).

https://doi.org/10.1016/j.jeconom.2023.105647
Received 6 July 2022; Received in revised form 16 December 2023; Accepted 16 December 2023
Available online 9 January 2024

Similar (weak) identification problems occur in many other econometric models, e.g. instrumental variable models, nonlinear regression models and many others, see Staiger and Stock (1997), Stock and Wright (2000) and Andrews and Mikusheva (2015) for some examples. The key difference between this existing literature and the non-Gaussian LSEM is that, in the latter, the parameters responsible for the possible identification failure are density functions, i.e. infinite dimensional parameters. Therefore, whilst conceptually the identification problem is the same, providing robust inferential methods requires a new approach which is capable of handling identification failure caused by infinite dimensional nuisance parameters.

To this extent, this paper develops a new approach for conducting inference in LSEMs that is inspired by the weak identification robust methods developed in econometrics (e.g. Stock and Wright, 2000; Kleibergen, 2005; Andrews and Mikusheva, 2015) and the general semiparametric statistical theory that is discussed in Bickel et al. (1998) and van der Vaart (2002). In brief, we treat the LSEM as a semiparametric model, where the densities of the independent structural shocks are treated non-parametrically, and we construct confidence bands for the possibly unidentified structural parameters of interest by inverting semiparametric score tests. The approach efficiently exploits non-Gaussianity when it is present in the data and yields confidence bands which do not asymptotically under-cover under sequences of densities that are local (in a $\sqrt{n}$ neighborhood) to the true density. Moreover, the test is easy to implement and the critical values accompanying the test statistic are standard chi-squared.

The effective score test that we propose is the semi-parametric analog of the Neyman–Rao test (e.g. Neyman, 1979; Hall and Mathiason, 1990). In the conventional Neyman–Rao test the scores for the parameter of interest are orthogonalized with respect to the scores for the *finite dimensional* nuisance parameters. In our setting the nuisance parameter includes the densities of the shocks, i.e. an *infinite dimensional* parameter. While such nuisance functions result in the orthogonal projection being more technically demanding to derive, the main idea of Neyman (1979) continues to apply.

Formally, we show that the semi-parametric score test is locally robust in the sense that its null rejection probability is no greater than the nominal level under parameter sequences that can be described by local deviations from the true parameters which satisfy the null hypothesis. In particular, the null rejection probability of the test is controlled for sequences of densities that converge at a $\sqrt{n}$ rate to the Gaussian density, a point of identification failure. These sequences are the natural counterpart in our setting to the "weak identification asymptotics" as found in, for example, Staiger and Stock (1997), Stock and Wright (2000), Moreira (2003), Kleibergen (2005), Andrews and Mikusheva (2015). Moreover, they are those considered in the theory of Kaji (2021) who studies estimation in weakly identified semi-parametric models.[1] In addition, we show that under strong identification, which requires all errors to be (sufficiently) non-Gaussian, the score test is semi-parametrically efficient in the sense that it attains various local asymptotic power bounds for testing scalar or vector valued parameters (cf. Choi et al., 1996).

We evaluate the finite sample performance of the semiparametric score test in a large simulation study. We find that the null rejection probability of our test remains close to the nominal level for all distributions considered, including those which are "close" to the Gaussian distribution and the Gaussian distribution itself. In contrast, tests that are based on the sampling variation of (pseudo)-maximum likelihood or GMM estimators often substantially over-reject in weakly non-Gaussian settings. Further, for moderate sample sizes the power of the semiparametric test is comparable to the parametric score test that relies on knowing the functional form of the density. When the parametric density of the (pseudo)-maximum likelihood score test is misspecified the semi-parametric test is always found to be preferable.

To showcase the empirical value of our methodology we adopt the score test to construct confidence bands for the effect of education on wages. To do so, we consider a special case of the LSEM model: the linear instrumental variable (IV) model. We show that the presence of independent non-Gaussian errors allows to (i) strengthen identification for the case where the instrument is assumed exogenous and (ii) test and correct for endogenous instrumental variables. We emphasize that our theory allows for, and is locally robust to, weak instruments.

For the model specification and data considered in Card (1995) we find that inverting the semi-parametric score test gives the shortest confidence intervals for the returns to education which are, for instance, shorter when compared to confidence intervals based on the Anderson and Rubin (1949) statistic. Also, when we relax the instrument exogeneity assumption and use non-Gaussianity to identify the returns to education, we find that (i) the assumption that the proximity to college instrument is exogenous cannot be rejected and (ii) the confidence interval for the returns to education remains precisely estimated. In contrast, using alternative but non-efficient methods we find considerably larger confidence sets when relaxing the instrument exogeneity assumption.

In general, this paper highlights the problem of weak non-Gaussianity and provides a solution in the setting of i.i.d. linear simultaneous equations models. We point out that similar non-Gaussian identification approaches have been adopted in other settings and it is likely that weak non-Gaussianity continues to cause inference problems for standard MLE and GMM methods in these settings. Prominent examples include (i) structural VAR(MA) models (Lanne and Lütkepohl, 2010; Moneta et al., 2013; Lanne et al., 2017; Maxand, 2018; Gouriéroux et al., 2019; Tank et al., 2019; Herwartz, 2019; Herwartz et al., 2019; Bekaert et al., 2020, 2021; Lanne and Luoto, 2021; Guay, 2021; Sims, 2021; Moneta and Pallante, 2022; Gabriele and Sentana, 2023; Velasco, 2023; Davis and Ng, 2022; Drautzburg and Wright, 2023), (ii) measurement error models (e.g. Reiersøl, 1950; Kapteyn and Wansbeek, 1983; Dagenais and Dagenais, 1997; Erickson and Whited, 2000, 2002; Bonhomme and Robin, 2009), and (iii) triangular systems (e.g. Lewbel et al., 2023). In future work we aim to extend our semi-parametric inference approach to cover these more general settings. The supplementary material that accompanies this paper provides a step in this direction by considering a class of nonlinear simultaneous equations models.

[1] See also Andrews and Mikusheva (2022) who study weakly identified GMM models using the same type of local sequences.

Further, as mentioned above, this paper shows that the proposed semi-parametric score test has null rejection probability asymptotically bounded by the nominal level under weak identification asymptotics, i.e. under parameter sequences representing local deviations from the true parameters (which satisfy the null hypothesis).[2] In order to prove asymptotic size control of the score test, one would need to show that the same holds under *all* nuisance parameter sequences which satisfy the null hypothesis, i.e. also those that are outside a $\sqrt{n}$ neighborhood of the true parameters. For models where identification strength is determined by the value of a finite dimensional parameter, results of this nature are provided by, inter alia, Andrews and Cheng (2012), Andrews and Cheng (2013) and Andrews et al. (2020). Whether results of this nature can be extended to cover cases (such as that considered in the present paper) where identification strength is determined by an infinite dimensional parameter is an important topic for future research; it remains an open question whether asymptotic size control can be achieved in a meaningful way in such a setting, i.e. without unreasonably restricting the parameter space and/or equipping it with a very strong metric.

The remainder of this paper is organized as follows. In the next section we provide a simple example that illustrates the identification problem and intuitively discusses our solution. Section 3 presents the main LSEM model and provides the implementation details for the effective score test. Section 4 discusses the main theoretical results including the required assumptions. Sections 5 and 6 summarize the results from the simulation and empirical studies. Section 7 concludes. Any references to sections, equations, lemmas etc. which start with "S" refer to the supplementary material.

2. Illustrative example

In this section we use a simple example to illustrate: (i) the identification problem in LSEMs, (ii) why conventional inference methods suffer from size distortions when the structural shocks have densities close to Gaussian and (iii) how our proposed approach aims to circumvent such distortions.

The identification problem

Consider the simple bi-variate model

$$Y_i = A^{-1}\epsilon_i \,, \qquad i = 1, \dots, n \,, \tag{1}$$

where Y_i is a vector of observable variables, A is a rotation matrix (i.e. $A^{-1} = A'$ and $\det(A) = 1$) and ϵ_i is a vector with independent structural shocks ϵ_{ik}, for $k = 1, 2$, that have mean zero, unit variance and common density η. For concreteness, we will parameterize the rotation matrix as follows

$$A = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix} \,, \tag{2}$$

where $\alpha \in [0, 2\pi)$ and we let α_0 denote the true parameter.[3]

Model (1) has two parameters: the parameter of interest α and the infinite dimensional nuisance parameter η. Suppose for now that η is known and let the log likelihood function for Y_i be denoted by $\ell_\alpha(\cdot)$. The parameter α is locally identified if the expected score of $\ell_\alpha(Y_i)$ with respect to α is non-zero for all $\alpha \neq \alpha_0$ in a neighborhood of α_0.

Whether local identification occurs depends crucially on η. To illustrate, consider the case where η is equal to the Gaussian density. Since ϵ_i is normalized we have

$$\mathbb{E}\ell_\alpha(Y_i) \propto -\frac{1}{2}\mathbb{E}(AY_i)'(AY_i) = -1 \,,$$

and hence the expected loglikelihood takes the same value irrespective of α. This is plotted in the top left panel of Fig. 1, where we show the expected likelihood $\mathbb{E}\ell_\alpha(Y_i)$ as a function of α with $\alpha_0 = \pi$ as the true parameter (an arbitrary choice). This illustrates the standard identification problem in linear simultaneous equations models: without additional identifying restrictions, the impact effects of the structural shocks are not identifiable when the structural shocks follow a Gaussian distribution.

The other plots in Fig. 1 show that this is no longer the case when we move away from the Gaussian distribution. In each case the expected gradient becomes non-zero at values $\alpha \neq \alpha_0$ in a neighborhood of α_0, i.e. local identification occurs. While for the (standardized) Student's t distribution with five degrees of freedom (i.e. $t(5)$) the change in the value of the expected likelihood is substantial it is easy to see that for more modest deviations from Gaussianity (e.g. $t(15)$) the difference is less pronounced. Further, note that non-Gaussian densities do not ensure α is globally identified, instead identification is only up to permutation and sign of the shocks.

Finite sample size distortions

In population α is always locally identified when all but one component of η is non-Gaussian (e.g. Comon, 1994, Theorem 11), but this is not sufficient for good performance of standard testing procedures in finite samples. In particular, if the structural shocks are too close to Gaussian, the available identifying information may be small relative to the sampling variability. Standard asymptotic approximations are not reliable in this setting and, as a result, testing procedures based on these approximations may fail to provide reliable inference.

[2] Such sequences have also been used to model weak identification in semi-parametric models in Kaji (2021), Andrews and Mikusheva (2022).

[3] Note that in our general framework we will not restrict A to be a rotation matrix nor η to be common. This example is chosen for exposition purposes only and corresponds to the case where the variance of Y_i is normalized to unity.

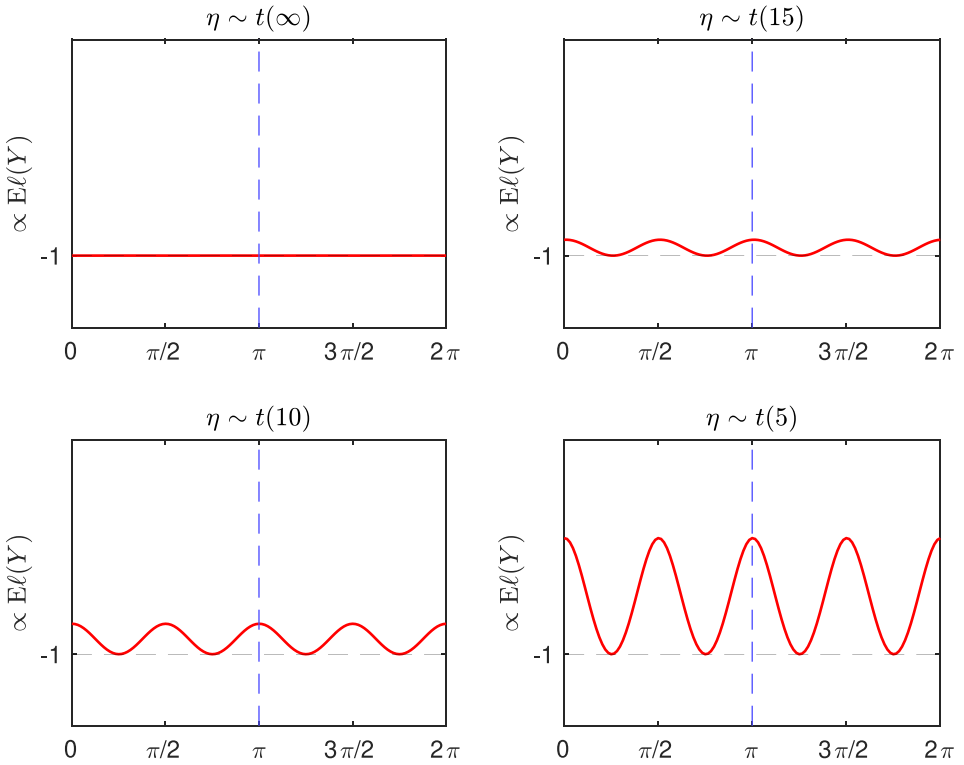

Fig. 1. (Weak) Non-Gaussian identification.
Notes: In the figure we show the expected log likelihood (red line) as a function of $\alpha \in [0, 2\pi)$. The true value is $\alpha_0 = \pi$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

To illustrate how the density η affects standard inference methods in finite sample, we draw 5000 samples $\{Y_i\}_{i=1}^n$ from model (1) for different η's using different sample sizes $n = 250, 500, 750$. Fig. 2 shows the finite sample distribution of the t-statistic for the hypothesis $H_0 : \alpha = \alpha_0$, with $\alpha_0 = \pi$, based on the maximum likelihood estimator under the assumption that η is known. The blue dashed lines show the $\mathcal{N}(0, 1)$ density that corresponds to the usual limit of the t-statistic. As can clearly be seen in this figure, the quality of the approximation provided by the standard Gaussian depends crucially on the underlying density, η. For a given sample size, the approximation deteriorates substantially the closer η is to a standard Gaussian density.

This deterioration results in poor size control of standard tests. Table 1 shows the empirical rejection frequencies for three standard tests in the same setting: Wald (W), likelihood ratio (LR) and Lagrange multiplier (LM) (or score) tests, all computed under the assumption that η is known. The empirical rejection frequencies correspond to the test for $H_0 : \alpha = \alpha_0$ with nominal level $a = 0.05$, where the critical values are based on the standard χ^2_1 asymptotic approximation.

We find that the Wald test is severely size distorted for η close to Gaussian; in view of the poor quality of asymptotic approximation depicted in Fig. 2 this is not surprising. As η gets closer to Gaussianity, the likelihood ratio test starts to under-reject as when α is poorly identified the likelihood values are very similar. Both of these tests are based on estimates of α and, in weakly identified settings, such estimates will be inaccurate.

In contrast, the score test (LM) shows correct size as it fixes $\alpha = \alpha_0$ under the null and α does not need to be (well) identified for this test to be correctly sized. Intuitively, with α fixed and η known there are no further unknown elements in the scores and the remaining uncertainty is due to sampling variation. This observation provides the first building block for the test we will construct: it will be a score type test which fixes $\alpha = \alpha_0$ under the null.

Towards a semi-parametric score test

In practice, η will be unknown. To build up to our semi-parametric approach, consider first the case where η is known up to a finite dimensional parameter vector, say ν. For example ν may include the degrees of freedom of the Student's t distribution.

For such cases Neyman (1979) proposed a convenient extension of the standard score test, that amounts to first orthogonalizing the scores for α with respect to the scores for ν and then computing a quadratic form of the score statistic. To illustrate let $\dot{\ell}(Y_i) = (\dot{\ell}_\alpha(Y_i), \dot{\ell}_\nu(Y_i))'$, $\dot{\ell}_\alpha(Y_i) = \nabla_\alpha \ell_{\alpha,\nu}(Y_i)$, $\dot{\ell}_\nu(Y_i) = \nabla_\beta \ell_{\alpha,\nu}(Y_i)$ and $\hat{I} = \frac{1}{n}\sum_{i=1}^n \dot{\ell}(Y_i)\dot{\ell}(Y_i)'$, denote the score and information matrix for α and ν. The Neyman–Rao score test statistic is given by

$$S = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{\kappa}(Y_i)\right)' \hat{I}^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{\kappa}(Y_i)\right),$$

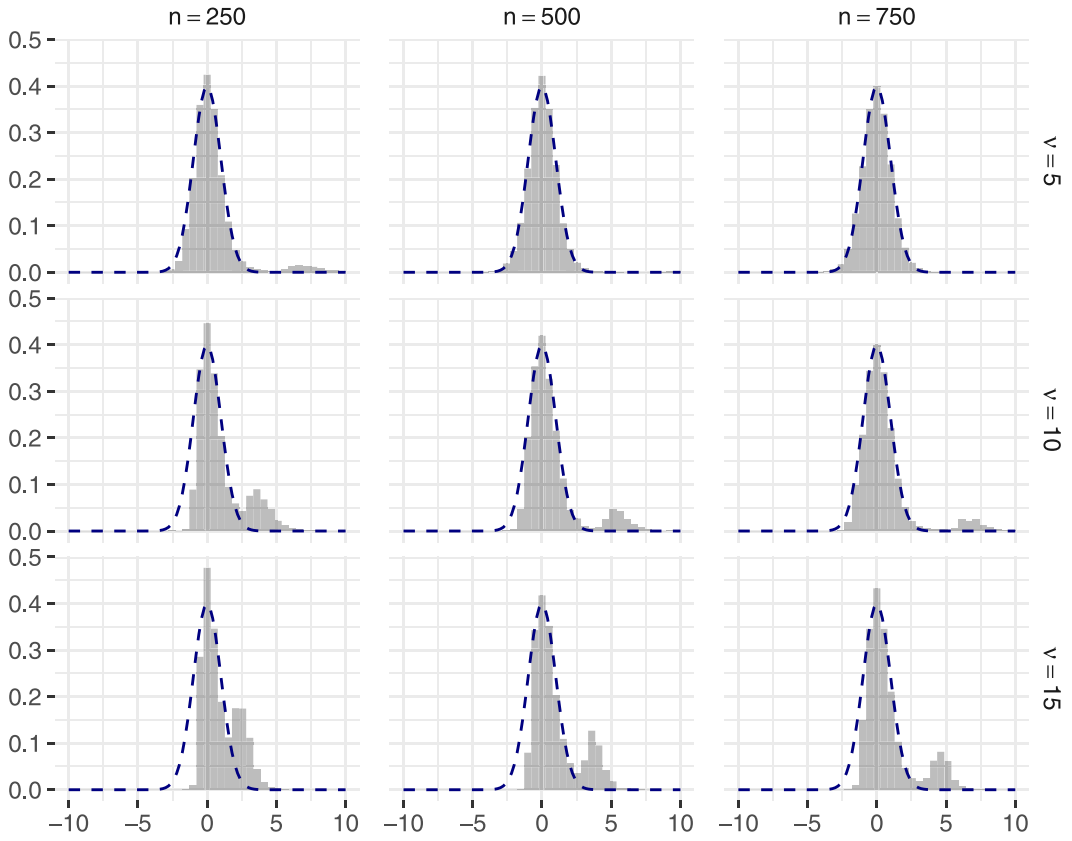

Fig. 2. Poor asymptotic approximation close to Gaussianity.
Notes: In the figure we show the finite sample distribution of the t-statistic based on the maximum likelihood estimator of α (the true value is $\alpha_0 = \pi$) for different sample sizes (n) and different degrees of freedom (ν) in the (standardized) t distribution, all based on 5000 replications. Letting $\hat{\alpha}$ be the ML estimator and α_0 the null hypothesis value of α, the t-statistic used is $t = \sqrt{n}(\hat{\alpha} - \alpha_0) \times \sqrt{\hat{I}}$, with $\hat{I}$ the usual outer product of gradients (OPG) estimator of the (Fisher) information: $\hat{I} = \frac{1}{n}\sum_{i=1}^{n} \hat{\ell}_{\hat{\alpha}}(Y_i)^2$, with $\hat{\ell}_{\alpha} = \nabla_{\alpha}\ell_{\alpha}$.

Table 1
Rejection frequencies for ML tests close to Gaussianity.

	t(15)			t(10)			t(5)		
n	W	LM	LR	W	LM	LR	W	LM	LR
250	25.26	4.42	3.74	20.56	4.24	4.04	8.88	4.84	4.08
500	21.76	4.54	4.52	13.10	4.38	3.60	6.38	4.42	4.92
750	17.12	4.96	3.94	9.90	4.88	3.42	6.12	5.28	5.64

Notes: The table shows the empirical rejection frequencies for the three maximum likelihood tests, under the assumption that η is known and based on 5000 Monte Carlo replications for the baseline model $Y_i = R'\epsilon_i$. The test has nominal level $a = 0.05$.

with

$$\hat{\kappa}(Y_i) = \hat{\ell}_{\alpha}(Y_i) - \hat{I}_{\alpha\nu}\hat{I}_{\nu\nu}^{-1}\hat{\ell}_{\nu}(Y_i) \qquad \text{and} \qquad \hat{\mathcal{I}} = \hat{I}_{\alpha\alpha} - \hat{I}_{\alpha\nu}\hat{I}_{\nu\nu}^{-1}\hat{I}_{\nu\alpha}\ ,$$

where $\hat{I}_{\cdot\cdot}$ denote the corresponding blocks of $\hat{I}$.[4] The (estimated) orthogonalized scores $\hat{\kappa}(\cdot)$ are often referred to as the (estimates of the) effective scores and $\hat{\mathcal{I}}$ is the corresponding (estimate of the) effective information matrix.

This score statistic is usually evaluated as $\alpha = \alpha_0$ and some $\sqrt{n}$ consistent estimate for ν. Whenever such an estimate exists, S will converge to a standard χ^2 limit under the null provided that $\hat{\mathcal{I}}$ is invertible.[5] In such cases, tests based on S retain correct

[4] This is numerically equivalent to the "usual" score test when the nuisance parameter ν is estimated by (restricted) maximum likelihood under the null hypothesis (Kocherlakota and Kocherlakota, 1991).

[5] In our results below we allow $\hat{\mathcal{I}}$ to be singular and rely on an eigenvalue truncated generalized inverse, see also Andrews (1987), Lütkepohl and Burda (1997) and Andrews and Guggenberger (2019).

size regardless of whether or not α is well identified making them attractive for settings where identification failure due to finite dimensional nuisance parameters is a concern (e.g. Andrews and Mikusheva, 2015).

Unfortunately, there are two distinct problems that may arise in the solution sketched above. First and most practically relevant, modeling the deviations from the Gaussian density in a parametric manner may result in biases and/or lower power whenever the true density lies outside of the parametric class considered. Second, parametric deviations from the Gaussian density as captured by ν generally nest the Gaussian distribution. In many such cases the information matrix associated to ν, i.e. $I_{\nu\nu}$, becomes singular when the true density is Gaussian. Sometimes this problem can be circumvented by re-parametrizing ν, e.g. parameterize $\tilde{\nu} = \nu^{-1}$ for the degrees of freedom of the Student's t or for a skewed-normal one can adopt the centered parametrization of Azzalini and Capitanio (2014, Section 3.1.4). However, for other examples, such as mixtures of normals, there are no available transformations that prevent the information matrix from becoming singular under Gaussianity. That is, ν itself becomes unidentified (Rothenberg, 1971, Theorem 1) and consistent estimators of ν do not exist.

We note that these problems interact as solving the identification problem for ν can be done by adopting a pseudo maximum likelihood approach that fixes ν at some reasonable value (e.g. Gouriéroux et al., 2017), but this immediately implies that the true likelihood may be far away from the fixed pseudo likelihood, resulting in a test with little power.

In the present paper, we do not assume that the parametric form of η is known up to a finite dimensional parameter vector but instead treat η non-parametrically. To avoid the creation of additional identification problems we rely on B-spline estimators to non-parametrically estimate the aspect of η which is necessary to implement our procedure: the log density score of η (i.e. the logarithmic derivative of η). Unlike the finite dimensional parameters ν discussed above, the log density score does not suffer from identification problems at Gaussianity.

Despite such changes, the underlying logic of our approach is similar to that sketched above. We first orthogonalize the score for α with respect to the scores for η and obtain a semi-parametric analog of the conventional Neyman–Rao score test. This requires technical adjustments as the scores with respect to η need to be defined differently and the projection with respect to η scores requires more care. For this we follow the semi-parametric literature as outlined in Bickel et al. (1998) and van der Vaart (2002).

3. Locally robust inference for LSEMs

In this section we propose a semi-parametric score test for testing parameters in a general class of linear simultaneous equations models. We first introduce the model class and give some motivating examples. Thereafter, we present a heuristic derivation for the score test and the exact implementation details. All theoretical properties including the main assumptions are deferred to the next section.

3.1. General model, objectives and examples

We consider the linear simultaneous equations model for a random sample of K endogenous variables Y_i, d exogenous variables $X_i = (1, \tilde{X}_i')'$ and K independent structural shocks ϵ_i, which have mean zero and unit variance. Specifically, we have

$$Y_i = BX_i + A^{-1}\epsilon_i \,, \qquad i = 1, \dots, n \,, \tag{3}$$

where we observe $W_i = (Y_i', X_i')'$ and the matrices B and A^{-1} map the explanatory variables and the structural shocks to the endogenous variables. The density functions of the components of $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{iK})'$ are denoted by $(\eta_1, \dots, \eta_K)$ and the density of $\tilde{X}_i$ is given by η_0. We set $\eta = (\eta_0, \eta_1, \dots, \eta_K)$.

As illustrated in the previous section, depending on the shapes of $\eta_1, \dots, \eta_K$ we may not be able to identify all parameters in A. To model this we let $A = A(\alpha, \sigma)$, where $A(\alpha, \sigma)$ is a function of both (i) the parameters α which may suffer from identification failure and (ii) the well-identified parameters in σ. We let $\alpha \in \mathcal{A} \subset \mathbb{R}^{L_\alpha}$ and set $\beta = (b, \sigma) \in \mathcal{B} \subset \mathbb{R}^{L_b} \times \mathbb{R}^{L_\sigma} = \mathbb{R}^{L_\beta}$, with $b = \text{vec}(B)$.

In this paper we leave the parametrization of $A = A(\alpha, \sigma)$ largely unspecified. In Assumption 1 we state the formal requirements and subsequently provide examples that can be adopted within our general framework. We stress that the dimensions of α and σ are fixed, as is the dimension of Y_i. As such our framework does not deal with high dimensional LSEMs. A special case of model (3) is obtained when setting $B = 0$ for which the model reduces to the baseline model for independent components analysis (e.g. Hyvärinen et al., 2001). Further, after pre-whitening the residuals we obtain the model (1) from the illustrative example.

The general LSEM (3) depends on the following parameters

$$\theta = (\gamma, \eta) \,, \quad \text{with} \quad \gamma = (\alpha, \beta) \quad \text{and} \quad \beta = (b, \sigma) \,, \tag{4}$$

where $\gamma \in \Gamma = \mathcal{A} \times \mathcal{B}$ summarizes all finite dimensional parameters, including the possibly weakly identified α and the well identified β, and η includes the infinite dimensional parameters, i.e. the densities of the shocks for which the parameter space will be formalized below.

We are interested in testing the possibly weakly identified parameters α. To do so, we consider the hypothesis

$$H_0 : \alpha = \alpha_0 \qquad \text{against} \qquad H_1 : \alpha \neq \alpha_0 \,. \tag{5}$$

Tests for such H_0 can then be inverted to yield confidence sets for α. A related set-up is found in Risk et al. (2019) and Jin et al. (2019) who assume that the structural shocks can be separated into *exactly* Gaussian and non-Gaussian shocks. We do not impose such structure, but we note that if indeed shocks can be separated in this way our approach will remain valid, but likely less efficient when compared to Risk et al. (2019).

Parameterizing the LSEM. In practice, we can adopt different parametrizations for modeling $A = A(\alpha, \sigma)$ in (3). A general requirement is that A is non-singular and that it is sufficiently smooth with respect to α and σ. The following assumption formalizes these conditions.

Assumption 1. Define the partial derivative matrices $D_{\alpha,l} = \partial A(\alpha,\sigma)/\partial \alpha_l$, for $l = 1, \dots, L_\alpha$, and $D_{\sigma,m} = \partial A(\alpha,\sigma)/\partial \sigma_m$, for $m = 1, \dots, L_\sigma$. Further, for each $i, j \in \{1, \dots, K\}$, $l \in \{1, \dots, L_\alpha\}$ and $m \in \{1, \dots, L_\sigma\}$ define $\zeta^{\alpha}_{l,k,j} := [D_{\alpha,l}]_{k\bullet} A^{-1}_{\bullet j}$ and $\zeta^{\sigma}_{m,k,j} := [D_{\sigma,m}]_{k\bullet} A^{-1}_{\bullet j}$, where the notation $M_{\bullet j}$ or $M_{j\bullet}$ denotes the jth column or row (respectively) of a matrix M. We assume that for all $(\alpha, \beta) \in \mathcal{A} \times \mathcal{B}$

1. $A(\alpha, \sigma)$ is non-singular
2. $(\alpha, \sigma) \mapsto A(\alpha, \sigma)$ is continuously differentiable
3. $(\alpha, \sigma) \mapsto \zeta^{\alpha}_{l,k,j}(\alpha, \sigma)$ and $(\alpha, \sigma) \mapsto \zeta^{\sigma}_{m,k,j}(\alpha, \sigma)$ are locally Lipschitz continuous for all j, k, l, m

The following examples illustrate some possible parametrizations that are of practical interest and satisfy the smoothness assumptions.

Example 1 (*Supply and Demand*)**.** Following Working (1927)'s canonical analysis of supply and demand curves let Y^s_{i1} and Y^d_{i1} denote the quantity demanded and supplied of some good with price Y_{i2}. In equilibrium we have $Y^d_{i1} = Y^s_{i1}$ and a simple model (omitting covariates for convenience) is given by

$$Y_{i1} = \alpha_1 Y_{i2} + \sigma_1 \epsilon_{i1} \qquad \text{(demand)}$$
$$Y_{i1} = \sigma_3 Y_{i2} + \sigma_2 \epsilon_{i2} \qquad \text{(supply)}$$

where ϵ_{i1} and ϵ_{i2} are independent demand and supply shocks. We can accommodate this model in our general framework by letting $\sigma = (\sigma_1, \sigma_2, \sigma_3)$ and defining the mapping $A(\alpha, \sigma)$ according to

$$A(\alpha,\sigma) = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}^{-1} \begin{bmatrix} 1 & -\alpha_1 \\ 1 & -\sigma_3 \end{bmatrix}.$$

Note that even with non-Gaussian errors, which we do not assume, the matrix $A(\alpha, \sigma)$ is only identifiable up to post multiplication by DP, where P is a permutation matrix and D a diagonal matrix with elements ± 1 on the main diagonal (e.g. Comon, 1994, Theorem 11). In applications we could impose sign restrictions to select the permutation that is of economic interest. For instance, here we could impose $\alpha_1 \leq 0$ and $\sigma_3 \geq 0$ to ensure that the demand curve is downward sloping and the supply curve is upward sloping, as well as $\sigma_1, \sigma_2 > 0$ to ensure that the scales are positive. As such we would only test values for α_1 in (5) that satisfy the sign restrictions.

Example 2 (*Instruments*)**.** In the context of the previous example, a common identification approach is based on using instrumental variables. Suppose that Y_{i3} is an instrument that correlates with the supply shock but is believed to be uncorrelated with demand, an assumption that we would like to test. After re-defining the errors and parameters we can write the model as

$$Y_{i1} = \alpha_1 Y_{i2} + \sigma_1 \epsilon_{i1}$$
$$Y_{i1} = \sigma_4 Y_{i2} + \sigma_5 Y_{i,3} + \sigma_2 \epsilon_{i2}$$
$$Y_{i3} = \alpha_2 \epsilon_{i,1} + \sigma_3 \epsilon_{i3}$$

where $\alpha_2 = 0$ implies that the instrument is exogenous and $\sigma_5 \neq 0$ implies that the instrument is relevant. We have $Y_i = (Y_{i1}, Y_{i2}, Y_{i3})'$, $\epsilon_i = (\epsilon_{i1}, \epsilon_{i2}, \epsilon_{i3})'$ and

$$A(\alpha,\sigma) = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ \alpha_2 & 0 & \sigma_3 \end{bmatrix}^{-1} \begin{bmatrix} 1 & -\alpha_1 & 0 \\ 1 & -\sigma_4 & -\sigma_5 \\ 0 & 0 & 1 \end{bmatrix}.$$

With this parametrization we have several options. First, assuming that the instruments are exogenous we set $\alpha_2 = 0$, and use the non-Gaussian errors to provide additional identifying information for α_1. This could be of use when instruments are weak. Second, we can relax the instrument exogeneity assumption and jointly test $\alpha = (\alpha_1, \alpha_2)$. This allows to simultaneously assess the slope of the demand curve and the exogeneity of the instrument. If the instruments are irrelevant and the errors are Gaussian we will not be able to reject any value.

Example 3 (*Rotation Matrix*)**.** As in Gouriéroux et al. (2017) we can set $A(\alpha,\sigma)^{-1} = \Sigma^{1/2}(\sigma) R(\alpha)'$, where $\Sigma^{1/2}(\sigma)$ is lower triangular with parameters σ and $R(\alpha)$ is a rotation matrix. In this setting we have $\sigma = \text{vech}(\Sigma^{1/2})$ and α parametrizes R using the trigonometric transformation, the Cayley transformation or the exponential transformation of a skew-symmetric matrix (e.g. Gouriéroux et al., 2017; Magnus et al., 2021).

These examples highlight different options for parametrizing $A(\alpha, \sigma)$. In Examples 1 and 2 the parameter α_1 has a direct economic interpretation after an economically interesting permutation has been selected using either sign restrictions or the instrumental variable. In Example 2 the parameter α_2 has an econometrically interesting interpretation: if it is non-zero, the instrument is not exogenous. In Example 3 the parameters α do not have a direct structural interpretation, but this specification corresponds to a common choice in the ICA literature (e.g. Hyvärinen et al., 2001; Gouriéroux et al., 2017).

3.2. Effective score test for LSEMs

Next, we provide a step by step implementation guide for the semi-parametric score test that aims to test $H_0 = \alpha = \alpha_0$. We postpone the theoretical justification of the test to the next section.

Effective score and information matrix. As intuitively explained in the simple example of Section 2, the proposed score test for the null hypothesis $H_0 : \alpha = \alpha_0$ is of the Neyman–Rao type, which relies on the effective scores for the parameters of interest α. Loosely speaking these scores are defined as the projection of the score function for α on the orthogonal complement of the space spanned by the score functions for the nuisance parameters (β, η) (e.g. Choi et al., 1996; Bickel et al., 1998; Newey, 1990; van der Vaart, 2002).

In the case of interest here, where the nuisance parameter contains both finite (β) and infinite-dimensional (η) components, the effective score function can be calculated in two steps: (1) compute the projection of the score for $\gamma = (\alpha, \beta)$ on the orthocomplement of the space spanned by the score functions for η, and (2) partition the resulting object into the components corresponding to α and β and project the former onto the orthocomplement of the latter.

For step (1) we follow Amari and Cardoso (1997) and Chen and Bickel (2006) who derive this projection for a special case of the LSEM (3) where B is known to be 0, i.e. the ICA model. The log likelihood contribution for observation W_i from model (3) is given by

$$\ell_\theta(W_i) = \log|A| + \sum_{k=1}^{K} \log \eta_k(A_{k\bullet} V_i) + \log \eta_0(\tilde{X}_i) \,,$$

where $V_i = Y_i - BX_i$.[6] The scores (i.e. partial derivatives of ℓ_θ) with respect to the components of α, σ and b are denoted by $\dot{\ell}_{\theta,\alpha_l} = \nabla_{\alpha_l} \ell_\theta$, $\dot{\ell}_{\theta,\sigma_l} = \nabla_{\sigma_l} \ell_\theta$ and $\dot{\ell}_{\theta,b_l} = \nabla_{b_l} \ell_\theta$. The effective scores are obtained by projecting $\dot{\ell}_{\theta,\alpha_l}, \dot{\ell}_{\theta,\sigma_l}$ and $\dot{\ell}_{\theta,b_l}$ on the orthocomplement of the space spanned by the score functions for η:[7]

$$\mathcal{T} = \left\{ w \mapsto h_0(\tilde{x}) + \sum_{k=1}^{K} h_k(A_{k\bullet}(y - Bx)) : h = (h_0, h_1, \dots, h_K) \in H = \prod_{k=0}^{K} H_k \right\} \tag{6}$$

where $x = (1, \tilde{x}')'$, $w = (y', x')'$. H_0 is the space of bounded functions $h_0 : \mathbb{R}^{d-1} \to \mathbb{R}$ which satisfy $\mathbb{E} h_0(\tilde{X}_i) = 0$. For $k = 1, \dots, K$, H_k is the space of functions $h_k : \mathbb{R} \to \mathbb{R}$ which are bounded and continuously differentiable with bounded derivative and satisfy $\mathbb{E}[h_k(\epsilon_{i,k})] = \mathbb{E}[\epsilon_{i,k} h_k(\epsilon_{i,k})] = \mathbb{E}[\kappa(\epsilon_{i,k}) h_k(\epsilon_{i,k})] = 0$, with $\kappa(z) = 1 - z^2$. The set $\mathcal{T}$ is the collection of scores corresponding to $\eta = (\eta_0, \eta_1, \dots, \eta_K)$: the densities of $\tilde{X}_i$ and $\epsilon_{i1}, \dots, \epsilon_{iK}$, see Lemma S1 in the supplementary material for a formal statement.

Intuitively, each $h_k \in H_k$ is restricted such that $\eta_k(1 + t h_k)$ is a density function and satisfies the conditions imposed by the model (for all small enough t). For instance, for $k = 1, \dots, K$, the restrictions on h_k ensure that $\epsilon_{ik} = A_{k\bullet} V_i$ remains mean zero and with variance one under the density $\eta_k(1 + t h_k)$. The elements of the set $\mathcal{T}$ are obtained by taking the derivative of the log likelihood evaluated at $\theta_t = (\gamma, \eta_0(1 + t h_0), \dots, \eta_K(1 + t h_K))$ with respect to t and evaluating this at $t = 0$, for a given $h = (h_0, \dots, h_K) \in H$; see van der Vaart (1998, Section 25.3) for a general discussion.

The effective scores are then defined as $\tilde{\ell}_{\theta,\alpha_l} = \dot{\ell}_{\theta,\alpha_l} - \Pi \dot{\ell}_{\theta,\alpha_l}$, $\tilde{\ell}_{\theta,\sigma_l} = \dot{\ell}_{\theta,\sigma_l} - \Pi \dot{\ell}_{\theta,\sigma_l}$ and $\tilde{\ell}_{\theta,b_l} = \dot{\ell}_{\theta,b_l} - \Pi \dot{\ell}_{\theta,b_l}$, where Π denotes the projection on $\mathrm{cl}\,\mathcal{T}$, the closure of $\mathcal{T}$. We compute these projections analytically to obtain

$$\tilde{\ell}_{\theta,\alpha_l}(W_i) = \sum_{k=1}^{K} \sum_{j=1, j\neq k}^{K} \zeta^{\alpha}_{l,k,j} \phi_k(A_{k\bullet} V_i) A_{j\bullet} V_i + \sum_{k=1}^{K} \zeta^{\alpha}_{l,k,k} \left[\tau_{k,1} A_{k\bullet} V_i + \tau_{k,2} \kappa(A_{k\bullet} V_i) \right]$$

$$\tilde{\ell}_{\theta,\sigma_l}(W_i) = \sum_{k=1}^{K} \sum_{j=1, j\neq k}^{K} \zeta^{\sigma}_{l,k,j} \phi_k(A_{k\bullet} V_i) A_{j\bullet} V_i + \sum_{k=1}^{K} \zeta^{\sigma}_{l,k,k} \left[\tau_{k,1} A_{k\bullet} V_i + \tau_{k,2} \kappa(A_{k\bullet} V_i) \right]$$

$$\tilde{\ell}_{\theta,b_l}(W_i) = \sum_{k=1}^{K} [-A_{k\bullet} D_{b,l}] \left[(X_i - \mathbb{E} X_i) \phi_k(A_{k\bullet} V_i) - \mathbb{E} X_i \left(\varsigma_{k,1} A_{k\bullet} V_i + \varsigma_{k,2} \kappa(A_{k\bullet} V_i) \right) \right]$$

where $\zeta^{\alpha}_{l,k,j}$ and $\zeta^{\sigma}_{l,k,j}$ are defined in Assumption 1, $D_{b,l} = \partial B / \partial b_l$ and $\phi_k(x) = \partial \log \eta_k(x) / \partial x$. Further,

$$\tau_k = M_k^{-1} \begin{pmatrix} 0 \\ -2 \end{pmatrix}, \quad \varsigma_k = M_k^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \text{where } M_k = \begin{pmatrix} 1 & \mathbb{E}_\theta(A_{k\bullet} V_i)^3 \\ \mathbb{E}_\theta(A_{k\bullet} V_i)^3 & \mathbb{E}_\theta(A_{k\bullet} V_i)^4 - 1 \end{pmatrix}.$$

The derivations that lead to these expressions are given the supplementary material where Lemma S3 provides the formal statement. The expressions show that the effective scores depend on the log density scores ϕ_k, i.e. the non-parametric part stemming from η_k, and the third and fourth moments of the errors $A_{k\bullet} V_i$ via the vectors τ_k and ς_k, for $k = 1, \dots, K$.

For step (2) we will project the effective scores for α on the space spanned by the effective scores for $\beta = (b, \sigma)$. Since the latter space is finite dimensional this projection takes a standard form. First, we collect and partition the effective scores as follows

$$\tilde{\ell}_\theta(W_i) = \begin{bmatrix} \tilde{\ell}_{\theta,\alpha}(W_i) \\ \tilde{\ell}_{\theta,\beta}(W_i) \end{bmatrix} \quad \text{and} \quad \tilde{\ell}_{\theta,\beta}(W_i) = \begin{bmatrix} \tilde{\ell}_{\theta,\sigma}(W_i) \\ \tilde{\ell}_{\theta,b}(W_i) \end{bmatrix},$$

[6] Throughout the main text the dependence of e.g. V_i, A, $D_{x,l}$ and $\zeta^{x}_{l,k,j}$, with $x \in \{\alpha, \sigma\}$, on (parts of) γ is left implicit.

[7] Each score function lies in $L_2(P_\theta)$, which is the Hilbert space under consideration here.

where $\tilde{\ell}_{\theta,\alpha} = (\tilde{\ell}_{\theta,\alpha_1}, \dots, \tilde{\ell}_{\theta,\alpha_{L_\alpha}})'$, $\tilde{\ell}_{\theta,\sigma} = (\tilde{\ell}_{\theta,\sigma_1}, \dots, \tilde{\ell}_{\theta,\sigma_{L_\sigma}})'$ and $\tilde{\ell}_{\theta,b} = (\tilde{\ell}_{\theta,b_1}, \dots, \tilde{\ell}_{\theta,b_{L_b}})'$ are the $L_\alpha \times 1$, $L_\sigma \times 1$ and $L_b \times 1$ vectors that collect the effective score functions. With this notation we define the effective information matrix by

$$\tilde{I}_\theta = \mathbb{E}\tilde{\ell}_\theta(W_i)\tilde{\ell}_\theta'(W_i) \qquad \text{with partitioning} \quad \tilde{I}_\theta = \begin{pmatrix} \tilde{I}_{\theta,\alpha\alpha} & \tilde{I}_{\theta,\alpha\beta} \\ \tilde{I}_{\theta,\beta\alpha} & \tilde{I}_{\theta,\beta\beta} \end{pmatrix} .$$

The effective score function for α with respect to β and η can now be computed by the second projection (e.g. Bickel et al., 1998, p. 74)

$$\tilde{\kappa}_\theta(W_i) = \tilde{\ell}_{\theta,\alpha}(W_i) - \tilde{I}_{\theta,\alpha\beta}\tilde{I}_{\theta,\beta\beta}^{-1}\tilde{\ell}_{\theta,\beta}(W_i) . \tag{7}$$

The corresponding effective information matrix is given by

$$\tilde{I}_\theta = \tilde{I}_{\theta,\alpha\alpha} - \tilde{I}_{\theta,\alpha\beta}\tilde{I}_{\theta,\beta\beta}^{-1}\tilde{I}_{\theta,\beta\alpha} . \tag{8}$$

We note that the effective score function $\tilde{\kappa}_\theta(W_i)$ and the effective information matrix $\tilde{I}_\theta$ can be evaluated at any parameters $\theta = (\alpha, \beta, \eta)$.

Effective score and information matrix estimation. The effective scores and information depend on unknown nuisance parameters, such as the log density scores ϕ_k and the moment vectors τ_k and ζ_k. To implement the score test we replace these parameters by appropriate estimates. As we show in the supplementary material, consistent estimators for $\tilde{\ell}_\theta(W_i)$ are

$$\hat{\ell}_\gamma(W_i) = \begin{bmatrix} \hat{\ell}_{\gamma,\alpha}(W_i) \\ \hat{\ell}_{\gamma,\beta}(W_i) \end{bmatrix} \qquad \text{and} \qquad \hat{\ell}_{\gamma,\beta}(W_i) = \begin{bmatrix} \hat{\ell}_{\gamma,\sigma}(W_i) \\ \hat{\ell}_{\theta,b}(W_i) \end{bmatrix} ,$$

where the components are given by

$$\begin{aligned}
\hat{\ell}_{\gamma,\alpha_l}(W_i) &= \sum_{j,k=1, j\neq k}^{K} \zeta^\alpha_{l,k,j}\hat{\phi}_k(A_k{\bullet}V_i)A_j{\bullet}V_i + \sum_{k=1}^{K}\zeta^\alpha_{l,k,k}\left[\hat{\tau}_{k,1}A_k{\bullet}V_i + \hat{\tau}_{k,2}\kappa(A_k{\bullet}V_i)\right] \\
\hat{\ell}_{\gamma,\sigma_l}(W_i) &= \sum_{j,k=1, j\neq k}^{K} \zeta^\sigma_{l,k,j}\hat{\phi}_k(A_k{\bullet}V_i)A_j{\bullet}V_i + \sum_{k=1}^{K}\zeta^\sigma_{l,k,k}\left[\hat{\tau}_{k,1}A_k{\bullet}V_i + \hat{\tau}_{k,2}\kappa(A_k{\bullet}V_i)\right] , \\
\hat{\ell}_{\gamma,b_l}(W_i) &= \sum_{k=1}^{K}[-A_k{\bullet}D_{b,l}][(X_i - \bar{X})\hat{\phi}_k(A_k{\bullet}V_i) - \bar{X}(\hat{\varsigma}_{k,1}A_k{\bullet}V_i + \hat{\varsigma}_{k,2}\kappa(A_k{\bullet}V_i))]
\end{aligned} \tag{9}$$

with $\bar{X} = n^{-1}\sum_{i=1}^n X_i$. The coefficients $\hat{\tau}_k = (\hat{\tau}_{k,1}, \hat{\tau}_{k,2})'$ and $\hat{\varsigma}_k = (\hat{\varsigma}_{k,1}, \hat{\varsigma}_{k,2})'$ are given, for $k = 1, \dots, K$, by

$$\hat{\tau}_k = \hat{M}_k^{-1}\begin{pmatrix} 0 \\ -2 \end{pmatrix}, \quad \hat{\varsigma}_k = \hat{M}_k^{-1}\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \hat{M}_k = \begin{pmatrix} 1 & \frac{1}{n}\sum_{i=1}^n (A_k{\bullet}V_i)^3 \\ \frac{1}{n}\sum_{i=1}^n (A_k{\bullet}V_i)^3 & \frac{1}{n}\sum_{i=1}^n (A_k{\bullet}V_i)^4 - 1 \end{pmatrix} . \tag{10}$$

The estimates for the effective scores can be evaluated at any $\gamma = (\alpha, \beta)$, but do not depend on ϕ_k, τ_k, ς_k or $\mathbb{E}X_i$ as these components have been replaced by estimators $\hat{\phi}_k$, $\hat{\tau}_k$, $\hat{\varsigma}_k$ and $\bar{X}$. These estimators may depend on γ and the index n, though this is left implicit in the notation.

Density score estimation. The log density score estimates $\hat{\phi}_k(\cdot)$ needed for computing (9) can be obtained in different ways and our preferred approach is based on using B-splines as in Jin (1992) and Chen and Bickel (2006). We can define these estimates by

$$\hat{\phi}_k(z) = \hat{\psi}_k' b_k(z) , \text{ with } \quad \hat{\psi}_k = -\left[\sum_{i=1}^n b_k(A_k{\bullet}V_i)b_k(A_k{\bullet}V_i)'\right]^{-1}\sum_{i=1}^n c_k(A_k{\bullet}V_i) , \tag{11}$$

where z is the argument of the function, e.g. $z = A_k{\bullet}V_i$ in (9), $b_k(z) = (b_{k,1}(z), \dots, b_{k,\mathsf{B}_k}(z))'$ is a collection of B_k cubic B-splines and $c_k(z) = (c_{k,1}(z), \dots, c_{k,\mathsf{B}_k}(z))'$ are their derivatives: $c_{k,i}(z) = \frac{\mathrm{d}b_{k,i}(z)}{\mathrm{d}z}$ for each $i = 1, \dots, \mathsf{B}_k$, see de Boor (2001) for more details on B-splines.[8] In practice we rely on equally spaced knots with upper and lower end points taken to be the 95th and 5th percentile of the samples $\{\epsilon_i\}_{i=1}^n$ adjusted by $\log(\log(n))$. We use $\mathsf{B}_k = 6$ splines in our main simulations below and investigate the sensitivity of this choice.

Given the estimates of the effective scores we estimate the effective information matrix, which is the variance matrix of the effective score function, as

$$\hat{I}_\gamma = \frac{1}{n}\sum_{i=1}^n \hat{\ell}_\gamma(W_i)\hat{\ell}_\gamma(W_i)' \qquad \text{with partitioning} \quad \hat{I}_\gamma = \begin{bmatrix} \hat{I}_{\gamma,\alpha\alpha} & \hat{I}_{\gamma,\alpha\beta} \\ \hat{I}_{\gamma,\beta\alpha} & \hat{I}_{\gamma,\beta\beta} \end{bmatrix} . \tag{12}$$

With these estimates we can compute the estimates for the effective score of α with respect to β and η, i.e. $\tilde{\kappa}_\theta(W_i)$ as defined in (7), and the corresponding information matrix (8).

$$\hat{\kappa}_\gamma(W_i) = \hat{\ell}_{\gamma,\alpha}(W_i) - \hat{I}_{\gamma,\alpha\beta}\hat{I}_{\gamma,\beta\beta}^{-1}\hat{\ell}_{\gamma,\beta}(W_i) \qquad \text{and} \qquad \hat{I}_\gamma = \hat{I}_{\gamma,\alpha\alpha} - \hat{I}_{\gamma,\alpha\beta}\hat{I}_{\gamma,\beta\beta}^{-1}\hat{I}_{\gamma,\beta\alpha} . \tag{13}$$

[8] Further details as required for the construction in this paper are given in Section S6 in the supplementary material. For the asymptotic theory, B_k will be required to (slowly) diverge with n. In the main text we omit the dependence of B_k and b_k on n in the notation.

Importantly, $\tilde{\mathcal{I}}_\theta$ may not be positive definite in our setting. For instance, when the densities η_k correspond to the Gaussian density, $\tilde{\mathcal{I}}_\theta$ is singular, see the discussion preceding Lemma S19 in the supplementary material.

Semi-parametric score statistic. With $\hat{\kappa}_\gamma$ and $\hat{\mathcal{I}}_\gamma$ we can define the semi-parametric score test statistic for the LSEM model as a function of $\gamma = (\alpha, \beta)$ and the observations W_i by

$$\hat{S}_\gamma = \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\kappa}_\gamma(W_i) \right)' \hat{\mathcal{I}}_\gamma^{t,\dagger} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\kappa}_\gamma(W_i) \right) , \tag{14}$$

where $\hat{\mathcal{I}}_\gamma^{t,\dagger}$ denotes the generalized inverse of the eigenvalue truncated effective information matrix $\hat{\mathcal{I}}_\gamma$ (cf. Lütkepohl and Burda, 1997). Formally,

$$\hat{\mathcal{I}}_\gamma^{t} = \hat{U} \hat{\Lambda}(v_n^{1/2}) \hat{U}' , \tag{15}$$

where $\hat{\Lambda}(v_n^{1/2})$ is a diagonal matrix with the $v_n^{1/2}$-truncated eigenvalues of $\hat{\mathcal{I}}_\gamma$ on the main diagonal and $\hat{U}$ is the matrix of corresponding orthonormal eigenvectors. To be specific, let $\{\hat{\lambda}_i\}_{i=1}^{L}$ denote the non-increasing eigenvalues of $\hat{\mathcal{I}}_\gamma$, then the (i,i)th element of $\hat{\Lambda}(v_n^{1/2})$ is given by $\hat{\lambda}_i \mathbf{1}(\hat{\lambda}_i \geq v_n^{1/2})$. We discuss the choice for the truncation parameter in more detail below.

Eqs. (9)–(15) define the semi-parametric score statistic for the LSEM model (3) for a given parameter vector $\gamma = (\alpha, \beta)$. To test the null hypothesis (5) we will evaluate this test statistic at $\alpha = \alpha_0$, i.e. fixing the possibly unidentified parameters under the null, and at $\hat{\beta}$, which can be any $\sqrt{n}$ consistent estimate for β. Let $\hat{\gamma} = (\alpha_0, \hat{\beta})$. In our simulations, we use ordinary least squares estimates for σ and $b = \text{vec}(B)$, or one-step efficient estimates following van der Vaart (2002, Section 7.2). In our theoretical section below we show that under suitable assumptions the score statistic will converge to a χ^2 limit. Specifically, we prove that under H_0 for any $a \in (0,1)$ we have

$$\lim_{n\to\infty} P(\hat{S}_{\hat{\gamma}} > c_n) \leq a , \tag{16}$$

where c_n is the $1-a$ quantile of the $\chi^2_{r_n}$ distribution with $r_n = \text{rank}(\hat{\mathcal{I}}^t_{\hat{\gamma}})$. Importantly, as we show in Section 4 this result does not rely on any assumptions regarding the shape of the densities η, i.e. we do not need to assume that η is non-Gaussian. Only conventional moment assumptions and some regularity conditions on the densities are required. The following algorithm summarizes the complete implementation.

Algorithm: Effective score test for LSEM

1. Obtain $\sqrt{n}$-consistent estimates $\hat{\beta} = (\hat{\sigma}, \hat{b})$, residuals $\hat{V}_i = Y_i - \hat{B} X_i$ and evaluate all quantities in steps **2–5** at $\hat{\gamma} = (\alpha_0, \hat{\beta})$;
2. For $k = 1, \dots, K$, compute $\hat{\phi}_k(\hat{A}_{k\cdot}\hat{V}_i)$ from (11) with $\hat{A} = A(\alpha_0, \hat{\sigma})$;
3. Compute the effective scores $\hat{\ell}_{\hat{\gamma}}(W_i)$ from (9) and the information matrix $\hat{I}_{\hat{\gamma}}$ from (12);
4. Compute $\hat{\kappa}_{\hat{\gamma}}(W_i)$ and $\hat{\mathcal{I}}_{\hat{\gamma}}$ from (13) and $\hat{\mathcal{I}}^t_{\hat{\gamma}}$ from (15) using truncation parameter $v_n^{1/2}$;
5. Compute the score statistic $\hat{S}_{\hat{\gamma}}$ from (14) and reject $H_0 : \alpha = \alpha_0$ if $\hat{S}_{\hat{\gamma}} > c_n$, where c_n is the $1-a$ quantile of the $\chi^2_{r_n}$ distribution with $r_n = \text{rank}(\hat{\mathcal{I}}^t_{\hat{\gamma}})$.

The truncation parameter $v_n^{1/2}$ in step 4 is a tuning parameter for which the theoretical requirements are formalized in Assumption 3 below. In practice, we recommend a small tuning parameter (e.g. less than $v_n^{1/2} = 10^{-5}$) as our simulations suggest that the null rejection probability is well controlled for any such choice.[9] In practice the simplest implementation is to use a pseudo inverse function directly which implicitly truncates at machine precision. Nevertheless we recommend that researchers applying the proposed approach explore the performance of different choices of $v_n^{1/2}$ in simulation experiments designed to replicate the application at hand.

The algorithm highlights that the computational cost for evaluating the semi-parametric score statistic $\hat{S}_{\hat{\gamma}}$ is modest; effectively one only needs to compute K B-spline regressions to obtain the log density scores. Importantly, this implies that the algorithm can often be implemented without relying on numerical optimization routines.[10] Confidence sets for α can be constructed by inverting the score statistic over a range of values for α_0.

For some parametrizations of $A(\alpha, \sigma)$, the parameter of economic interest could be a function of both α and σ, or more generally, a function of α and $\beta = (b, \sigma)$. In these settings, the algorithm can be used in combination with the Bonferroni approach discussed in Granziera et al. (2018) to construct confidence intervals for such functions. Intuitively, this approach amounts to constructing a confidence set for $f(\alpha_0, \hat{\beta})$ with confidence level q_2 for each fixed α_0 for which the score test does not reject at level q_1, with $q_1 + q_2 = a$. Then, taking the union over the constructed sets for $f(\alpha_0, \hat{\beta})$ yields a $1-a$ confidence set for $f(\alpha, \beta)$.

[9] See Section S8.1 in the supplementary material for simulation results with different truncation values.

[10] Numerical optimization may be necessary to compute $\hat{\beta}$, depending on the chosen parametrisation, but is not necessary beyond this.

4. Asymptotic theory

In this section we present our main theoretical results. We start by carefully spelling out the regularity conditions that are required. After this we discuss the properties of the test. We show that (i) under weak identification asymptotics, the null rejection probability of the test does not exceed its nominal level asymptotically and (ii) under strong identification it attains well known power bounds for various classes of tests.

4.1. Assumptions

We assume that we observe a random sample $\{W_i\}_{i=1}^n = \{(Y_i', X_i')'\}_{i=1}^n$ from model (3) where the underlying components satisfy the following.

Assumption 2. For $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{iK})'$ in model (3), each component ϵ_{ik} has a continuously differentiable root density (with respect to Lebesgue measure on $\mathbb{R}$). We write the density as η_k with log density score $\phi_k(x) = \partial \log \eta_k(x)/\partial x$. We assume that for all $k = 1, \dots, K$ and some $\delta > 0$

1. $\mathbb{E}\epsilon_{ik} = 0$, $\mathbb{E}\epsilon_{ik}^2 = 1$, $\mathbb{E}\epsilon_{ik}^{4+\delta} < \infty$, $\mathbb{E}(\epsilon_{ik}^4) - 1 > \mathbb{E}(\epsilon_{ik}^3)^2$, and $\mathbb{E}\phi_k^{4+\delta}(\epsilon_{ik}) < \infty$;
2. $\mathbb{E}\phi_k(\epsilon_{ik}) = 0$, $\mathbb{E}\phi_k(\epsilon_{ik})\epsilon_{ik} = -1$, $\mathbb{E}\phi_k(\epsilon_{ik})\epsilon_{ik}^2 = 0$ and $\mathbb{E}\phi_k(\epsilon_{ik})\epsilon_{ik}^3 = -3$;
3. ϵ_{ik} is independent of ϵ_{il} for all $k \neq l$;
4. $\eta_0 \in \mathscr{Z}$ is a density function (with respect to Lebesgue measure on $\mathbb{R}^{d-1}$) such that if $\tilde{X}_i \sim \eta_0$, then $\mathbb{E}\tilde{X}_i\tilde{X}_i'$ is positive definite and $\mathbb{E}[|\tilde{X}_{i,l}|^{4+\delta}] < \infty$ for all $l = 1, \dots, d-1$;
5. ϵ_i and $\tilde{X}_i$ are independent.

The first part normalizes the errors to have mean zero, variance one and finite four+δ moments,[11] hence ruling out heavy tailed errors.[12] Additionally, we require the log density scores $\phi_k(x) = \partial \log \eta_k(x)/\partial x$ evaluated at the errors to have finite four+δ moments. The second part simplifies the construction of the effective score functions. Whilst this may at first glance appear a strong condition, Lemma S20 in the supplementary material shows that if the first part holds, then a simple sufficient condition is that the tails of the densities η_k converge to zero at a polynomial rate.[13] The third part imposes that the components of ϵ_i are independent. Part four imposes some structure on $\tilde{X}_i$ that allows us to identify B; notably positive definite second moments and four+δ finite moments are required. Part five requires the explanatory variables and errors to be independent. This can be relaxed by requiring the moment assumptions in 2 to hold conditional on $\tilde{X}_i$. In this setup, our general theory as outlined in this section would continue to be valid though the resulting effective score function would take a different form.

Most important is what is *not* in Assumption 2: there is no condition that imposes that a certain number of components of ϵ_i have a (sufficiently) non-Gaussian distribution.

The third assumption that we impose is only required for the estimation of the log density scores $\phi_k(x) = \partial \log \eta_k(x)/\partial x$ using B-spline regressions and can be appropriately replaced when a different density score estimator is used.[14] For notation purposes, let $\Xi_{k,n}^L$ and $\Xi_{k,n}^U$ denote the lower and upper endpoints of the cubic B-splines for $\phi_k(x)$ for $k = 1, \dots, K$. In practice, we select these points as the lower 5th and upper 95th percentiles of the samples $\{A_{k\bullet}V_i\}_{i=1}^n$ adjusted by $\log\log n$, see the implementation Section 3.

Assumption 3. Let ν_n be such that $\nu_{n,p}^2 = o(\nu_n)$ with $p := \min\{1 + \delta/4, 2\}$ and $\nu_{n,p} = n^{(1-p)/p}$ if $p \in (1,2)$ or $\nu_{n,p} = n^{-1/2}\log(n)^{1/2+\rho}$, for some $\rho > 0$, if $p = 2$. Let $\phi_{k,n} := \phi_k \mathbf{1}_{[\Xi_{k,n}^L, \Xi_{k,n}^U]}$ and $\Delta_{k,n} := \Xi_{k,n}^U - \Xi_{k,n}^L$ and suppose that for all $k = 1, \dots, K$, $[\Xi_{k,n}^L, \Xi_{k,n}^U] \uparrow \bar{\Xi} \supset \text{supp}(\eta_k)$ and $\delta_{k,n} \downarrow 0$

(i) $P(\epsilon_{ik} \notin [\Xi_{k,n}^L, \Xi_{k,n}^U]) = o(\nu_n^2)$;
(ii) For some $\iota > 0$, $n^{-1}\Delta_{k,n}^{2+2\iota}\delta_{k,n}^{-(8+2\iota)} = o(\nu_n)$;
(iii) η_k is bounded ($\|\eta_k\|_\infty < \infty$) and differentiable, with a bounded derivative: $\|\eta_k'\|_\infty < \infty$;
(iv) For each n, $\phi_{k,n}$ is three-times continuously differentiable on $[\Xi_{k,n}^L, \Xi_{k,n}^U]$ and $\|\phi_{k,n}^{(3)}\|_\infty^2\delta_{k,n}^6 = o(\nu_n)$;[15]
(v) There are $c > 0$ and $N \in \mathbb{N}$ such that for $n \geq N$ we have $\inf_{t \in [\Xi_{k,n}^L, \Xi_{k,n}^U]} |\eta_k(t)| \geq c\delta_{k,n}$.

First, the assumption provides conditions on the truncation rate $\nu_n^{1/2}$ that is needed for the truncation of the eigenvalues in (15). This rate is split into two parts. The "slow" rate $n^{(1-p)/p}$ (for $p \in (1,2)$) is always sufficient given Assumption 2, but if ϵ_{ik} has finite eighth moments the faster rate applies.

[11] $\mathbb{E}(\epsilon_{ik}^4) - 1 \geq \mathbb{E}(\epsilon_{ik}^3)^2$ always holds; this is known as Pearson's inequality. See e.g. result 1 in Sen (2012). Assuming that $\mathbb{E}(\epsilon_{ik}^4) - 1 > \mathbb{E}(\epsilon_{ik}^3)^2$ rules out (only) cases where $1, \epsilon_{ik}$ and ϵ_{ik}^2 are linearly dependent when considered as elements of L_2. See e.g. Theorem 7.2.10 in Horn and Johnson (2013).

[12] Heavy tailed errors in ICA and SVAR models have recently been considered in Davis and Ng (2022) and Davis and Fernandes (2022), but an inferential theory remains to be developed.

[13] See Example S1 in the supplementary material for an explicit example of a density which satisfies the first part of the assumption but not the second.

[14] See Assumption S1 for conditions on any alternative density score estimator under which our Theorem 1 continues to hold.

[15] The differentiability and continuity requirements at the end-points are one-sided.

Part (i) imposes that the tails of ϵ_{ik} decay to zero sufficiently fast.[16] Part (ii) ensures that the number of knots does not grow to fast relative to the sample size (and the truncation rate). Part (iii) requires the density and its derivative to be bounded. Part (iv) requires the existence of the third derivatives of ϕ_k and that the rate of increase of the third derivative is not too great. Part (v) ensures that the density is bounded away from zero on $[\Xi^L_{k,n}, \Xi^U_{k,n}]$. Overall, these assumptions are similar to those adopted in Chen and Bickel (2006), with two key differences.[17] Firstly, Chen and Bickel (2006) require the conditions to hold for the functions $v \mapsto \phi_k(A_k \cdot v)$ (rather than ϕ_k), uniformly over shrinking balls (at rate $n^{-1/2}$) around A. In our setting we are only interested in testing as consistent estimation is ruled out by the possible lack of identification, hence we only require the conditions to hold for the functions ϕ_k. Secondly, unlike Chen and Bickel (2006), we require convergence at a rate ν_n which satisfies certain decay conditions. This is due to the fact that we may have a singular effective information matrix and in order to obtain a consistent estimate of the Moore – Penrose inverse of this matrix, we require knowledge of the rate of convergence of our estimator.

4.2. Main results

In this section we formally state our main results for the semi-parametric score test $\hat{S}_{\hat{\gamma}}$. First, instead of evaluating the score test at the $\sqrt{n}$-consistent estimates $\hat{\gamma} = (\alpha_0, \hat{\beta})$ we will evaluate the score test at its discretized version $\bar{\gamma} = (\alpha_0, \bar{\beta}_n)$. Formally, let $\mathsf{G}_n = n^{-1/2} C \mathbb{Z}^{L_\beta}$ for some $C > 0$ and define $\bar{\beta}_n$ as a new version of $\hat{\beta}$ that replaces its value with the closest point in G_n. Note that this changes each coordinate of $\hat{\beta}$ by a quantity which is at most $O(n^{-1/2})$, hence the $\sqrt{n}$-consistency is retained by discretization. Since the constant C can be chosen arbitrarily small this change has no practical relevance for the implementation of the test.

The advantage of relying on discretized estimates is that it simplifies the proof of the main result. Specifically, it removes the need to show uniform convergence between the effective scores evaluated at $\hat{\beta}$ and β. The discretization trick is due to Le Cam (1960) and is widely used in statistics, see the detailed discussion in Le Cam and Yang (2000, Section 6.3), or van der Vaart (1998, page 72).[18]

The following theorem provides the main result.[19]

Theorem 1. *Suppose that Assumptions 1–3 hold and that (α_0, β) is an interior point of $\mathcal{A} \times \mathcal{B}$. Let $r_n = \operatorname{rank}(\hat{\mathcal{I}}^t_{\bar{\gamma}})$ and denote by c_n the $1 - a$ quantile of the $\chi^2_{r_n}$ distribution, for any $a \in (0,1)$. Then for any sequence*

$$\theta_n = \left(\alpha_0, \beta + d_n/\sqrt{n}, \eta(1 + h_n/\sqrt{n})\right) \ , \quad d_n \in D^\star \ , \ h_n \in H^\star \ ,$$

with $D^\star$ a bounded subset of $\mathbb{R}^{L_\beta}$ and $H^\star$ a compact subset of H, we have

$$\limsup_{n\to\infty} P^n_{\theta_n}(\hat{S}_{\bar{\gamma}} > c_n) \leq a,$$

with inequality only if $\operatorname{rank}(\tilde{\mathcal{I}}_{\theta_0}) = 0$ where $\theta_0 = (\alpha_0, \beta, \eta)$. The notation $P^n_{\theta_n}$ indicates the n-fold product of the measure P_{θ_n}, i.e. the distribution of the data $W_1, \ldots, W_n$ under θ_n.

A detailed proof for Theorem 1 can be found the supplementary material Section S1. The theorem shows that the test is locally robust in that its null rejection probability is no greater than the nominal a under any local sequence θ_n (consistent with the null). Under such sequences, the densities of the structural shocks (i.e. ϵ_{ik}) may converge to the Gaussian density at a $\sqrt{n}$ rate, i.e. these are local-to-Gaussian sequences. Studying the behavior of tests under these local-to-Gaussian sequences is the natural counterpart (in the model we study) to studying the performance of tests under so-called "weak identification asymptotics", as has been considered in many settings (e.g. Staiger and Stock, 1997; Stock and Wright, 2000; Moreira, 2003; Kleibergen, 2005; Andrews and Mikusheva, 2015). The key difference in our setting is that the identification failure occurs due to the value of an *infinite* dimensional nuisance parameter.

This local robustness follows from the fact that the test statistic $\hat{S}_{\bar{\gamma}}$ is locally regular, i.e. it attains its limiting distribution (under the null) in a locally uniform manner. This property, in turn, follows from the orthogonalization with respect to (all of) the nuisance parameters in the definition of the effective score function.[20] This orthogonalization ensures that the test statistic is insensitive to small deviations in the nuisance parameters and therefore that its limiting distribution does not change when the limit is taken along sequences of local alternatives consistent with the null hypothesis.

The result of Theorem 1 can be also written as

$$\limsup_{n\to\infty} \sup_{\theta \in \Theta_{0,n}} P^n_\theta(\hat{S}_{\bar{\gamma}} > c_n) \leq a \ ,$$

where

$$\Theta_{0,n} = \{(\alpha_0, \beta + d/\sqrt{n}, \eta(1 + h/\sqrt{n})) : d \in D^\star, h \in H^\star\} \ .$$

[16] The required speed of decay is linked to the truncation rate.

[17] Cf. their conditions C3, C5–C7, p. 2834.

[18] It has also been adopted in econometrics, see Cattaneo et al. (2012) for instance.

[19] The set H which appears in the statement of Theorem 1 is defined in Section 3. See Eq. (6) and the paragraph following it.

[20] In conjunction with the ULAN property shown to hold in Lemma S2.

This formulation allows us to highlight a difference between our *local* uniformity result, which is over local sets $\Theta_{0,n}$, and a more demanding *global* uniformity result in which the supremum would be taken over $\Theta_0 = \{(\alpha_0, \beta, \eta) : \beta \in \mathcal{B}, \eta \in \mathcal{H}\}$. We emphasize that Theorem 1 does not establish such a result.[21]

Efficiency under strong identification. Importantly, the local robustness of the score test does not come at the expense of power loss under strong identification. In particular, the test $\varphi_n := \mathbf{1}\{\hat{S}_{\tilde{\gamma}} > c_n\}$ is semiparametrically efficient when $\tilde{I}_\theta$ is nonsingular.[22] Here we provide a brief heuristic discussion of this point; proofs that these power bounds are attained by φ_n can be found in Section S7 of the supplementary appendix.

For the parameters $\theta = (\alpha, \beta, \eta)$ we consider local alternatives of the type

$$\theta_n(q, d, h) = \left(\alpha + q/\sqrt{n},\ \beta + d/\sqrt{n},\ \eta(1 + h/\sqrt{n})\right) . \tag{17}$$

First suppose that α is scalar and $\tilde{I}_\theta > 0$. Then the asymptotic power of the proposed test is against the local alternatives in (17) is

$$\lim_{n\to\infty} P^n_{\theta_n(q,d,h)}\varphi_n = 1 - \Phi\left(z_{a/2} - \tilde{I}_\theta^{1/2} q\right) + 1 - \Phi\left(z_{a/2} + \tilde{I}_\theta^{1/2} q\right) , \tag{18}$$

where Φ the standard normal CDF and $z_{a/2}$ the $1 - a/2$ quantile of the $\mathcal{N}(0, 1)$. This coincides with the (local asymptotic) power bound for locally asymptotically unbiased two sided tests of $q = 0$ against $q \neq 0$ (cf. Theorem 2 in Choi et al. (1996)).[23,24]

If instead α is multidimensional and $\tilde{I}_\theta$ is positive definite, then the asymptotic power of the proposed test is against the local alternatives in (17) is

$$\lim_{n\to\infty} P^n_{\theta_n(q,d,h)}\varphi_n = 1 - \mathrm{P}\left(\chi^2_{L_\alpha}(q'\tilde{I}_\theta q) \leq c_a\right), \tag{19}$$

where $\chi^2_r(u)$ denotes a random variable with a non-central χ^2 distribution with r degrees of freedom and non-centrality parameter u and c_a is the $1-a$ quantile of the (central) $\chi^2_{L_\alpha}$ distribution. This coincides with the (local asymptotic) power bound for asymptotically rotation invariant tests as developed in Section 5 of Choi et al. (1996) (see their Theorem 3).[25,26]

These power bounds make φ_n attractive in scenarios where there is no explicit direction in which one want to maximize power. When such directions are given alternative test statistics, also based on the effective score function, can be considered (e.g. Bickel et al., 2006). Maximin optimality results which permit singular $\tilde{I}_\theta$ matrices can be found in Lee (2023) for related tests in general semi-parametric models.

5. Simulation results

In this section we study the finite sample properties of the semi-parametric score test $\hat{S}_{\tilde{\gamma}}$. We study the empirical rejection frequency of the test under different data generating processes and compare its performance to several alternatives that have been proposed in the literature. We first study the simple model of Section 2 after which we consider the general linear simultaneous equations model (3). The supplementary material provides additional results.

5.1. Baseline model

We start by drawing independent samples from model (1), which we restate for convenience

$$Y_i = A^{-1}\epsilon_i , \qquad i = 1, \dots, n .$$

We take Y_i to be $K \times 1$ and consider $K = 2, 3$ and $K = 5$. The sample size is taken as $n = 200, 500$ or $n = 1000$. We fix ϵ_{i1} to have a standard Gaussian density and consider different densities for ϵ_{ik}, with $k = 2, \dots, K$. The non-Gaussian densities are either Student's t or mixtures of normals taken from Marron and Wand (1992). Fig. 3 provides an overview.

[21] For models where identification failures are determined by a finite dimensional η, global uniformity conditions are derived in Andrews and Cheng (2012, 2013) and Andrews et al. (2020). For the case where η is infinite dimensional much work remains to be done.

[22] Nonsingularity may fail to hold when multiple components of ϵ_i are Gaussian; see Lemma S19.

[23] One can alternatively see this by approximating the infinite dimensional model by a sequence of finite-dimensional models for which the corresponding result is well known and then taking limits. Cf. the proof of Theorem 25.44 in van der Vaart (1998).

[24] That the sequence of tests $(\varphi_n)_{n\in\mathbb{N}}$ is itself locally asymptotically unbiased is clear from (18).

[25] That the sequence of tests $(\varphi_n)_{n\in\mathbb{N}}$ is itself asymptotically rotation invariant is clear from (19): the limiting power function is that of the test $\varphi(Z) := \mathbf{1}\{Z'Z > c_a\}$ for $Z \sim \mathcal{N}(\tilde{I}_\theta^{1/2} q, I)$. This test is rotation invariant since for any rotation matrix R and any $z \in \mathbb{R}^{L_\alpha}$ one has $\varphi(R'z) = \mathbf{1}\{z'RR'z > c_a\} = \mathbf{1}\{z'z > c_a\} = \varphi(z)$.

[26] Related, the asymptotic maximin power of φ_n against the alternatives in (17) is

$$\lim_{n\to\infty} \inf_{(q,d,h)\in K_u^\star} P^n_{\theta_n(q,d,h)}\phi_n = 1 - \mathrm{P}\left(\chi^2_{L_\alpha}(u) \leq c_a\right), \tag{20}$$

where $K_u^\star$ is any compact subset of

$$K_u := \left\{(q, d, h) \in \mathbb{R}^{L_\alpha} \times \mathbb{R}^{L_\beta} \times H : q'\tilde{I}_\theta q \geq u\right\} .$$

which also coincides with the (local asymptotic) maximin power bound (cf. the parametric case in Theorem 13.5.5 of Lehmann and Romano (2005)).

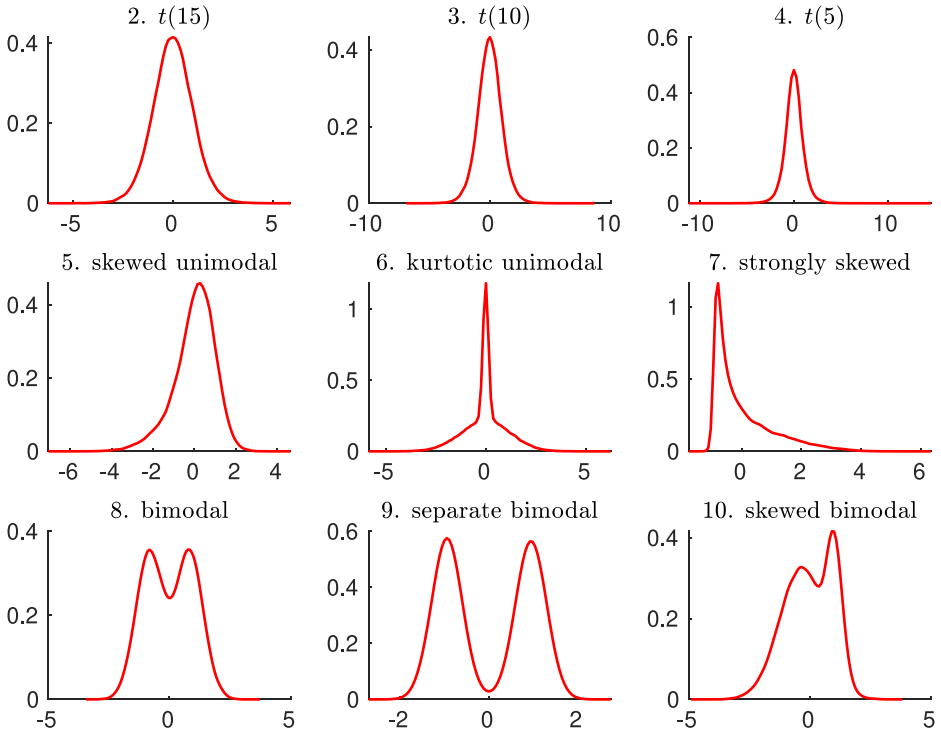

Fig. 3. Structural shock densities.
Notes: The plots show the different densities considered for simulating the structural shocks. Densities 2–4 are t-distributions normalized to have unit variance. Densities 5–10 (and their names) are mixtures of normals taken from Marron and Wand (1992); see their table 1 for the definitions. Density 1 is the standard Gaussian and omitted from the figure.

The matrix of interest A is taken as a rotation matrix and parametrized by the Cayley transformation of a skew-symmetric matrix (e.g. Gouriéroux et al., 2017):

$$A = A(\alpha) = (I - \Omega(\alpha))(I + \Omega(\alpha))^{-1} ,$$

where $\Omega(\alpha)$ is a skew-symmetric matrix (i.e. $\Omega(\alpha)' = -\Omega(\alpha)$) parameterized by α which we sample at random from $\alpha \sim N(0, I_{L_\alpha})$.

In this setting there are no additional nuisance parameters which allows us to concentrate on the consequences of weak non-Gaussianity on the semi-parametric score test and some alternative tests that have been proposed in the literature. In the simulation designs below we include additional finite dimensional nuisance parameters (i.e. $\beta = (b, \sigma)$) and investigate whether their inclusion alters the empirical rejection frequency of the test.

For each specification we simulate $S = 5000$ datasets and for each we compute the semi-parametric score statistic $\hat{S}_{\hat{\gamma}}$ as defined in Eq. (14) following the Algorithm given in Section 3.[27] We implement the log density score estimator (11) using $B = 4, 6$ or 8 cubic splines and truncate the effective information matrix at machine precision, i.e. $\nu_n^{1/2} = 10^{-308}$.

In Table 2 we show the empirical rejection frequencies under the null corresponding to the $S_{\hat{\gamma}}$ test with nominal level 0.05. The columns correspond to the different choices for the densities ϵ_{ik} for $k \geq 2$. The first column corresponds to the case where all densities are Gaussian and the expected likelihood takes the same value for all $\alpha \in \mathbb{R}^{L_\alpha}$, i.e. α is unidentified. Nonetheless, we find that the empirical rejection frequency of the score test is always close to the nominal level. This holds regardless of the sample size n, the dimension of the model K and the number of cubic splines B.

Second, when the densities for $k \geq 2$ are non-Gaussian the empirical rejection frequency remains approximately at the nominal level. Specifically, columns 2–4 show the results for the case where ϵ_{ik} follows a Student's t distribution with decreasing degrees of freedom ($\nu = 15, 10, 5$). No matter how close we get to the Gaussian density the empirical rejection frequency remains approximately at the nominal level. Columns 5–10 show that similar properties hold for a variety of mixture distributions. Even for complicated skewed bi-modal densities (e.g. columns 8–10) the $S_{\hat{\gamma}}$ test has empirical rejection frequency close to nominal regardless of the sample size.

[27] To be specific, since the model does not contain any finite dimensional nuisance parameters step 1 in the algorithm can be skipped and the score statistic is simply evaluated at α_0.

Table 2
Rejection frequencies $\hat{S}_{\tilde{\gamma}}$ test for Baseline model.

n	K	B	1	2	3	4	5	6	7	8	9	10
200	2	4	0.049	0.049	0.048	0.040	0.047	0.049	0.034	0.049	0.048	0.048
200	2	6	0.048	0.045	0.049	0.044	0.048	0.053	0.047	0.045	0.058	0.051
200	2	8	0.050	0.049	0.047	0.044	0.048	0.048	0.053	0.050	0.051	0.047
200	3	4	0.043	0.039	0.039	0.039	0.044	0.048	0.026	0.049	0.052	0.050
200	3	6	0.045	0.038	0.040	0.044	0.041	0.048	0.044	0.047	0.052	0.043
200	3	8	0.047	0.046	0.040	0.040	0.044	0.048	0.042	0.049	0.044	0.051
200	5	4	0.032	0.034	0.033	0.034	0.035	0.039	0.015	0.041	0.045	0.043
200	5	6	0.037	0.033	0.036	0.032	0.032	0.040	0.043	0.045	0.043	0.044
200	5	8	0.039	0.038	0.038	0.030	0.035	0.043	0.045	0.040	0.041	0.038
500	2	4	0.053	0.046	0.053	0.045	0.047	0.052	0.031	0.049	0.045	0.046
500	2	6	0.048	0.049	0.048	0.048	0.049	0.052	0.057	0.047	0.047	0.049
500	2	8	0.048	0.048	0.045	0.049	0.047	0.045	0.051	0.052	0.048	0.045
500	3	4	0.042	0.039	0.040	0.046	0.048	0.048	0.021	0.042	0.046	0.047
500	3	6	0.043	0.045	0.042	0.042	0.045	0.047	0.047	0.051	0.044	0.045
500	3	8	0.046	0.045	0.040	0.035	0.042	0.047	0.044	0.045	0.050	0.047
500	5	4	0.040	0.036	0.039	0.036	0.041	0.046	0.016	0.048	0.047	0.046
500	5	6	0.041	0.039	0.039	0.039	0.040	0.049	0.046	0.045	0.044	0.044
500	5	8	0.039	0.040	0.036	0.041	0.043	0.050	0.050	0.044	0.046	0.047
1000	2	4	0.042	0.052	0.040	0.055	0.047	0.052	0.046	0.052	0.046	0.048
1000	2	6	0.054	0.052	0.045	0.050	0.045	0.049	0.049	0.054	0.045	0.057
1000	2	8	0.047	0.048	0.048	0.047	0.048	0.052	0.050	0.048	0.055	0.052
1000	3	4	0.049	0.041	0.043	0.045	0.048	0.050	0.054	0.051	0.051	0.047
1000	3	6	0.048	0.044	0.038	0.040	0.050	0.047	0.046	0.049	0.051	0.045
1000	3	8	0.046	0.047	0.047	0.042	0.049	0.045	0.050	0.052	0.043	0.047
1000	5	4	0.038	0.035	0.038	0.047	0.041	0.044	0.050	0.046	0.047	0.048
1000	5	6	0.041	0.043	0.039	0.042	0.043	0.049	0.044	0.048	0.048	0.049
1000	5	8	0.042	0.042	0.038	0.039	0.048	0.050	0.049	0.047	0.045	0.049

Notes: The table shows the empirical rejection frequencies for the $S_{\tilde{\gamma}}$ test based on $S = 5000$ Monte Carlo replications for the baseline model $Y_i = A^{-1}\epsilon_i$. The test has nominal level $a = 0.05$. The columns denote the sample size n, the dimension of the model K, the number of B-splines B and the choice for densities ϵ_{ik}, for $k \geq 2$, where the numbers correspond to the different densities shown in Fig. 3.

Third, overall the number of cubic splines used has little influence on the results. A close inspection reveals that when the number of cubic splines is equal to four the test becomes mildly conservative for some densities, therefore we use $B = 6$ cubic splines in the remaining exercises.

Overall, the asymptotic approximation in Theorem 1 seems to provide a good approximation for the finite sample behavior of the semiparametric score test, at least for the densities shown in Fig. 3.

5.2. Comparison to alternative approaches

Next, we compare our semiparametric testing approach to different parametric approaches based on (psuedo) maximum likelihood and the generalized method of moments. We concentrate on evaluating different tests based on their empirical rejection frequency in the vicinity of Gaussianity.[28]

Alternative tests. Conceptually, there are two types of alternative tests that we consider: (i) tests that rely on estimates for α and (ii) tests that fix $\alpha = \alpha_0$ under the null. Clearly, from our intuitive discussion in Section 2 it follows that we expect tests that fix α under the null to perform relatively well.

In category (i) we consider the standard maximum likelihood Wald (W^{mle}) and likelihood ratio (LR^{mle}) tests based on the Student's t density for ϵ_k. For densities 2–4 in Fig. 3 these tests correspond to exact maximum likelihood tests, with the caveat that when the degrees of freedom increases the parameters α become weakly identified, or not-identified. For all other densities these tests are mis-specified.

In addition, we consider the pseudo-maximum likelihood Wald test (W^{pmle}) from Gouriéroux et al. (2017). This test is asymptotically valid for a broader range of true distribution functions and amount to fixing the functional form of the densities $\eta_1, \ldots, \eta_K$. We follow the implementation of Gouriéroux et al. (2017) and choose the Students t density with five degrees of freedom as the pseudo-likelihood and compute the Wald statistic based on this density.

Finally, we consider the recently developed GMM method of Lanne and Luoto (2021), which relies on higher order moments to identify the parameters α. We use $\mathbb{E}\epsilon_{ik}^2\epsilon_{ij} = 0$, $\mathbb{E}\epsilon_{ik}^3\epsilon_{ij} = 0$ and $\mathbb{E}\epsilon_{ik}^2\epsilon_{ij}^2 = 1$ as moment conditions for all $j \neq k$ and $j, k = 1, \ldots, K$. The

[28] The recent simulation studies of Herwartz et al. (2019) and Moneta and Pallante (2022) provide further simulation evidence for existing methods, also focusing on estimation accuracy.

Table 3
Rejection frequencies alternative tests for Baseline model.

Cat (i)	n	1	2	3	4	5	6	7	8	9	10
W^{mle}	200	0.179	0.149	0.139	0.127	0.113	0.059	0.097	0.152	0.125	0.171
	500	0.180	0.133	0.114	0.115	0.095	0.167	0.073	0.114	0.097	0.150
	1000	0.188	0.101	0.079	0.074	0.061	0.405	0.058	0.124	0.103	0.170
LR^{mle}	200	0.028	0.054	0.060	0.046	0.054	0.026	0.048	0.017	0.018	0.024
	500	0.043	0.056	0.068	0.054	0.065	0.023	0.053	0.016	0.017	0.024
	1000	0.049	0.065	0.063	0.061	0.053	0.031	0.051	0.022	0.018	0.025
W^{pmle}	200	0.375	0.211	0.198	0.086	0.141	0.058	0.105	0.495	0.998	0.467
	500	0.485	0.264	0.204	0.073	0.163	0.030	0.079	0.973	0.999	0.870
	1000	0.570	0.230	0.180	0.051	0.131	0.023	0.068	0.428	1.000	0.947
LR^{gmm}	200	0.413	0.411	0.425	0.441	0.290	0.379	0.120	0.216	0.086	0.232
	500	0.292	0.246	0.246	0.286	0.141	0.171	0.025	0.109	0.066	0.106
	1000	0.232	0.181	0.155	0.176	0.074	0.115	0.014	0.068	0.059	0.049
Cat (ii)	n	1	2	3	4	5	6	7	8	9	10
$\hat{S}_{\hat{\gamma}}$	200	0.051	0.047	0.048	0.040	0.049	0.049	0.047	0.048	0.050	0.044
	500	0.047	0.047	0.054	0.047	0.044	0.043	0.047	0.048	0.051	0.054
	1000	0.047	0.043	0.046	0.049	0.048	0.047	0.050	0.044	0.049	0.043
LM^{mle}	200	0.052	0.058	0.054	0.043	0.040	0.043	0.023	0.018	0.002	0.059
	500	0.056	0.052	0.052	0.042	0.046	0.047	0.028	0.017	0.001	0.062
	1000	0.062	0.052	0.050	0.049	0.039	0.040	0.029	0.016	0.002	0.052
LM^{plme}	200	0.049	0.045	0.049	0.035	0.038	0.046	0.030	0.041	0.042	0.042
	500	0.049	0.047	0.050	0.039	0.047	0.046	0.034	0.046	0.044	0.051
	1000	0.046	0.048	0.053	0.044	0.041	0.046	0.034	0.042	0.052	0.047
S^{gmm}	200	0.188	0.209	0.248	0.326	0.236	0.264	0.195	0.108	0.059	0.130
	500	0.094	0.105	0.123	0.223	0.116	0.133	0.103	0.057	0.028	0.064
	1000	0.061	0.070	0.081	0.162	0.069	0.078	0.054	0.031	0.019	0.035

Notes: The table shows the empirical rejection frequencies based on $S = 5000$ Monte Carlo replications for the baseline model $Y_i = A^{-1}\epsilon_i$, with $n = 500$ and $K = 2$. All tests have nominal level $a = 0.05$. The first column indicates the test the second the sample size. The remaining columns denote the choice for densities ϵ_{ik}, for $k \geq 2$, where the numbers correspond to the different densities shown in Fig. 3.

GMM likelihood ratio test is then computed as the rescaled difference between the unrestricted and restricted J-statistics, based on the 2-step GMM estimator (LR^{gmm}), see Lanne and Luoto (2021) for details.[29]

In category (ii) we consider tests which fix $\alpha = \alpha_0$ under the null. Specifically, we include the standard LM test (LM^{mle}) based on the Student's t density where the degrees of freedom parameter is estimated from the data. Second, we consider the pseudo-maximum likelihood version of the LM test (LM^{pmle}) based on Gouriéroux et al. (2017), which fixes the degrees of freedom at five. Finally, we consider the GMM-based identification robust S-statistic (S^{gmm}) of Stock and Wright (2000), which was recently considered in Drautzburg and Wright (2023) in the context of structural VAR models with non-Gaussian errors. We use the same moment conditions as considered in Drautzburg and Wright (2023) for the LM^{gmm} test.

Null rejection frequency comparison. We compare the empirical rejection frequencies of the different tests for the simulation designs described in Section 5.1. These are shown in Table 3 for the case where $K = 2$ and $n = 200, 500, 1000$. Overall we find, perhaps not surprisingly, that all tests in category (i) do not demonstrate the correct empirical rejection frequency when the true density is close to Gaussian nor when the corresponding method is based on a mis-specified model. This shows that tests based on estimates for α are generally unreliable. Tests in category (ii) overall demonstrate empirical rejection frequencies close to the nominal level.

More specifically, we find that the Wald tests (W^{mle} and W^{pmle}) tend to over-reject quite severely whilst the standard likelihood ratio test (LR^{mle}) tends to under-reject for most densities, especially in the vicinity of the Gaussian density, as ought to be expected given the earlier evidence in shown in Fig. 1.

Finally, the GMM likelihood ratio test (LR^{gmm}) also over-rejects, which confirms findings in Lanne and Luoto (2021) where the LR^{gmm} also over-rejects when the densities of the structural shocks are close to Gaussian.

In the second category the semi-parametric score test $\hat{S}_{\hat{\gamma}}$ (as proposed in this paper) and the pseudo maximum likelihood LM test (LM^{pmle}), inspired by Gouriéroux et al. (2017), both have near perfect empirical rejection frequencies across all densities. The standard LM test (LM^{mle}) also performs reasonably well, but when the functional form of the true densities is very different from the Student's t density (e.g. separate bi-modal, column 9) the test tends to under-reject.[30] Finally, the GMM based S test (S^{gmm}) tends to over-reject for small samples, but for large samples it generally shows correct size except for densities with moderately heavy tails such as the $t(5)$ density (column 4). In these cases the S^{gmm} over-rejects which can be understood when realizing that the GMM approach requires eight finite moments for inference when based on fourth-order moment restrictions. The $t(5)$ density does not have eight finite moments.

[29] Note that lower order moments are not required as the baseline model, $Y_i = A^{-1}\epsilon_i$ with A a rotation matrix, implies that the observations have mean zero and unit variance.

[30] Recall here that this test is based on a misspecified density.

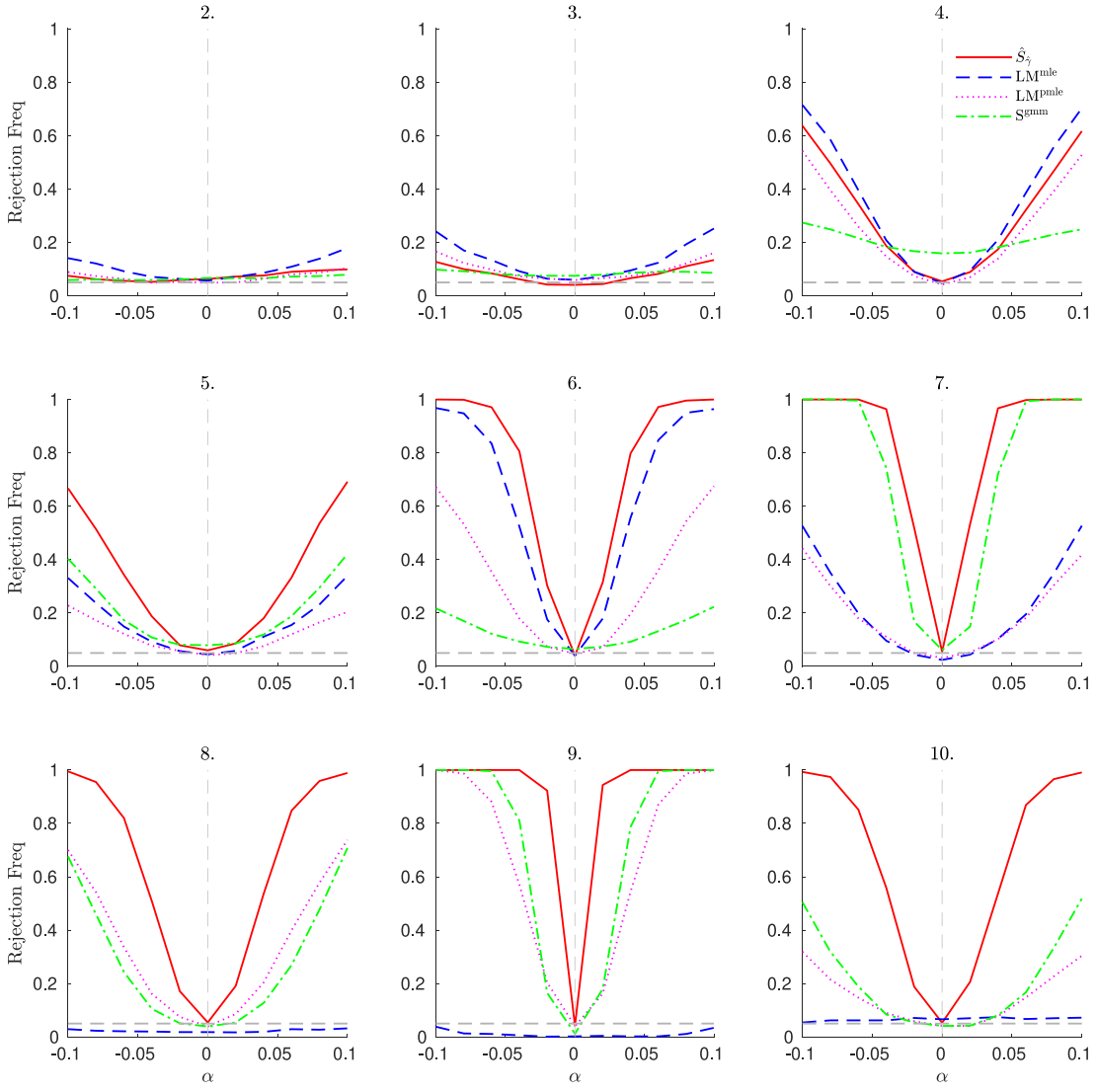

Fig. 4. Power comparison baseline model.
Notes: Empirical power curves for the baseline model with $k = 2$ and $n = 1000$. Each plot corresponds to the choice for densities ϵ_{ik}, for $k \geq 2$, where the numbers correspond to the different densities shown in Fig. 3. The solid red line corresponds to $S_{\tilde{\gamma}}$, the dashed blue line to LM^{mle}, the dotted pink line to LM^{pmle} and the dot-dashed green line to S^{gmm}. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In sum, we recommend avoiding statistics that are based on estimates for α as these are overall unreliable when the shock distributions are close to Gaussian. All tests that fix α under the null perform at least reasonably well.

Power comparison. We compare the power of all tests that fix α under the null, that is $\hat{S}_{\tilde{\gamma}}$, LM^{mle}, LM^{pmle} and S^{gmm}.

We consider the case where $K = 2$ and $n = 1000$.[31] In this setting α is a scalar parameter and we fixed the true value at 0 (an arbitrary choice). Fig. 4 shows the empirical rejection frequencies when we vary α around $\alpha = 0$. Each point on the curve is based on $S = 5000$ simulations.

Two main findings stand out. First, for the Student's t densities $t(15)$, $t(10)$ and $t(5)$ (panels 2–4) the standard LM test (LM^{mle}) shows the highest power. This is not surprising as for these data generating processes the LM^{mle} test is correctly specified and hence takes advantage of fitting the true densities using only a scalar parameter. That said, the semi-parametric score test ($\hat{S}_{\tilde{\gamma}}$) and the pseudo maximum likelihood LM test (LM^{pmle}) come reasonably close in terms of power.

[31] Power comparisons for different n can be found in the supplementary material.

Table 4
Rejection frequencies $\hat{S}_{\hat{\gamma}}$ test for LSEM - OLS $\hat{\beta}$.

n	K	d	1	2	3	4	5	6	7	8	9	10
200	2	2	0.050	0.054	0.049	0.049	0.038	0.030	0.038	0.043	0.057	0.046
200	2	3	0.049	0.054	0.054	0.048	0.046	0.059	0.042	0.035	0.029	0.052
200	3	2	0.056	0.058	0.050	0.062	0.059	0.031	0.018	0.038	0.047	0.050
200	3	3	0.063	0.054	0.057	0.065	0.060	0.025	0.023	0.051	0.058	0.049
200	5	2	0.098	0.104	0.109	0.142	0.094	0.051	0.064	0.054	0.023	0.057
200	5	3	0.116	0.116	0.131	0.155	0.103	0.039	0.029	0.061	0.026	0.072
500	2	2	0.049	0.050	0.039	0.042	0.041	0.027	0.029	0.036	0.026	0.029
500	2	3	0.048	0.041	0.047	0.047	0.037	0.029	0.024	0.034	0.050	0.051
500	3	2	0.051	0.051	0.048	0.040	0.037	0.028	0.029	0.038	0.022	0.039
500	3	3	0.048	0.050	0.047	0.051	0.053	0.028	0.048	0.041	0.037	0.036
500	5	2	0.071	0.078	0.068	0.081	0.049	0.023	0.060	0.042	0.039	0.038
500	5	3	0.067	0.068	0.080	0.085	0.063	0.022	0.045	0.049	0.027	0.051
1000	2	2	0.040	0.051	0.049	0.029	0.043	0.032	0.033	0.045	0.049	0.041
1000	2	3	0.048	0.044	0.040	0.040	0.040	0.030	0.038	0.046	0.030	0.044
1000	3	2	0.045	0.038	0.043	0.034	0.033	0.032	0.034	0.040	0.039	0.042
1000	3	3	0.044	0.045	0.043	0.036	0.030	0.032	0.035	0.040	0.024	0.034
1000	5	2	0.059	0.051	0.057	0.051	0.039	0.024	0.063	0.030	0.028	0.036
1000	5	3	0.057	0.058	0.056	0.050	0.035	0.018	0.046	0.036	0.029	0.040

Notes: The table shows the empirical rejection frequencies for the $S_{\hat{\gamma}}$ test based on $S = 5000$ Monte Carlo replications for the linear simultaneous equations model. The test has nominal level $a = 0.05$. The columns denote the sample size n, the dimension of the model K, the number of covariates d and the choice for densities ϵ_{ik}, for $k \geq 2$, where the numbers correspond to the different densities shown in Fig. 3. The $S_{\hat{\gamma}}$ test was implemented using $B = 6$ B-splines.

Second, for all other densities, i.e. different mixtures of normals in panels 5–10, the semi-parametric score test ($\hat{S}_{\hat{\gamma}}$) shows the highest power. Sometimes the difference with the other tests is not very large, but for instance for bi-modal densities (panels 8–10) the differences are substantial.

Overall, the good power of the $\hat{S}_{\hat{\gamma}}$ test corresponds to the theoretical finding that for non-singular information matrices the test is locally asymptotically uniformly most powerful in the class of (locally asymptotically) unbiased tests.

Besides the $\hat{S}_{\hat{\gamma}}$ test, we note that the pseudo maximum likelihood LM test and the GMM based S test shows quite promising power for most of the densities considered. Neither of these dominates the other. The caveat for the GMM test is that it is size-distorted for moderately heavy tails (panel 4).

5.3. Linear simultaneous equations model

Next, we discuss the simulation results for the general linear simultaneous equations model (3). The dimensions of the design are similar as above with the addition that we consider $d = 2, 3$ for the number of covariates. We now parametrize $A(\alpha, \sigma)^{-1} = \Sigma^{1/2}(\sigma)R(\alpha)$ as in Example 3, where $\Sigma^{1/2}$ is lower triangular and the rotation matrix R remains to be specified by the Cayley transform. The explanatory variables are drawn from the standard normal distribution.

The vector of finite dimensional nuisance parameters β now includes $\sigma = \text{vech}(\Sigma^{1/2})$ and $b = \text{vec}(B)$. Our main theoretical result in Theorem 1 permits any $\sqrt{n}$-consistent estimator of β. Obviously, ordinary least squares estimates are attractive for their simplicity, but given the non-normality of the structural shocks these estimators may be improved. Therefore we also consider estimating β by one-step-efficient estimates (e.g. van der Vaart, 2002, Section 7.2), which are easy to compute here since the effective score of β is computed anyway to construct the score test.

Similar to before, the first error $\epsilon_{i,1}$ follows a Gaussian distribution and the different densities from Fig. 3 are assigned to the other error terms. For each specification we simulate $S = 5000$ datasets and for each sample we compute the semi-parametric score statistic using the Algorithm in Section 3.

Null rejection frequency results. The empirical rejection frequencies are shown in Tables 4 and 5 for the OLS and one-step efficient estimates for β, respectively.

We find that for all densities the rejection frequencies of the $\hat{S}_{\hat{\gamma}}$ test are generally close to the nominal level. That said, there is more variation in the empirical rejection frequencies compared to Table 2, indicating that the estimation of the finite dimensional nuisance parameters does have consequences.

Starting with Table 4 where $\hat{\beta}$ is estimated by OLS. We find that the empirical rejection frequency of $\hat{S}_{\hat{\gamma}}$ is (approximately) the same regardless of how close the densities of ϵ_{ik} are to the Gaussian density. Specifically, moving from columns 1–4 (i.e. from Gaussian to $t(5)$) we see virtually no changes in the rejection frequencies. This holds for all specifications considered and highlights the main point of this paper: the semi-parametric score test yields reliable inference even when α is not, or poorly, identified.

Depending on the dimension of β we do find distortions in the empirical rejection frequencies for small sample sizes, most notably when $K = 5$ and $n = 200$. In this setting β is of dimension 20 or 25 depending on $d = 2, 3$, and we see that the test often over-rejects. This does not hold for all densities considered, but for Gaussian, Student's t and kurtotic unimodal densities the test over-rejects. When n increases this over-rejection vanishes.

Table 5
Rejection frequencies $\hat{S}_{\hat{\gamma}}$ test for LSEM – One-step $\hat{\beta}$.

n	K	d	1	2	3	4	5	6	7	8	9	10
200	2	2	0.067	0.080	0.068	0.081	0.070	0.031	0.054	0.056	0.061	0.051
200	2	3	0.068	0.074	0.076	0.072	0.066	0.071	0.057	0.047	0.026	0.061
200	3	2	0.095	0.106	0.104	0.120	0.090	0.041	0.026	0.059	0.036	0.061
200	3	3	0.099	0.103	0.105	0.114	0.098	0.037	0.028	0.071	0.035	0.064
200	5	2	0.187	0.226	0.247	0.264	0.178	0.063	0.040	0.072	0.020	0.068
200	5	3	0.212	0.238	0.262	0.289	0.193	0.064	0.049	0.089	0.036	0.088
500	2	2	0.062	0.062	0.068	0.067	0.057	0.034	0.049	0.041	0.021	0.037
500	2	3	0.059	0.064	0.071	0.069	0.056	0.031	0.019	0.046	0.031	0.051
500	3	2	0.078	0.078	0.081	0.079	0.066	0.026	0.024	0.047	0.021	0.045
500	3	3	0.076	0.081	0.091	0.088	0.068	0.025	0.029	0.050	0.042	0.042
500	5	2	0.112	0.149	0.158	0.181	0.097	0.036	0.035	0.060	0.030	0.044
500	5	3	0.129	0.151	0.168	0.180	0.101	0.033	0.023	0.069	0.031	0.058
1000	2	2	0.059	0.059	0.065	0.048	0.049	0.025	0.021	0.055	0.050	0.038
1000	2	3	0.060	0.060	0.060	0.068	0.057	0.038	0.052	0.050	0.027	0.051
1000	3	2	0.061	0.067	0.068	0.065	0.053	0.023	0.048	0.047	0.023	0.045
1000	3	3	0.064	0.066	0.072	0.070	0.054	0.040	0.016	0.047	0.022	0.041
1000	5	2	0.091	0.105	0.108	0.111	0.069	0.032	0.026	0.042	0.029	0.043
1000	5	3	0.085	0.102	0.120	0.103	0.065	0.026	0.020	0.047	0.026	0.050

Notes: The table shows the empirical rejection frequencies for the $\hat{S}_{\hat{\gamma}}$ test based on $S = 5000$ Monte Carlo replications for the linear simultaneous equations model (3). The test has nominal level $a = 0.05$. The columns denote the sample size n, the dimension of the observations K, the number of covariates d and the choice for densities ϵ_{ik}, for $k \geq 2$, where the numbers correspond to the different densities shown in Fig. 3. The $S_{\hat{\gamma}}$ test was implemented using $B = 6$ B-splines and using OLS estimates for β.

For the one-step efficient estimator for β the results are shown in Table 5. We find that on average the empirical rejection frequencies are larger when compared to the OLS estimator. Notably, when n is small over-rejection becomes more severe. Again, we find that this holds uniformly across all considered densities, i.e. the distortions do not depend on being close to Gaussianity, and the empirical rejection frequencies improve when n increases.

Power results. Next, we investigate the power of the $\hat{S}_{\hat{\gamma}}$ test for the LSEM model. We again consider the case where $K = 2$, $d = 2$ and $n = 1000$, which allows us to compare the results with those for the baseline model. The power curves are shown in Fig. 5 for both OLS and one-step estimates for β.

First, when comparing Fig. 5 to the case without nuisance parameters (i.e. Fig. 4) we find that the power of the test is reduced when we include nuisance parameters. Second, the power of the test using the one-step efficient estimates (dotted blue line) is higher when compared to the same test evaluated at OLS estimates. This holds for all densities considered.

Based on these results we recommend using OLS estimates for β when the sample size is small (e.g. $n = 200, 500$), but for larger sample sizes the one-step efficient estimates are preferable.

6. Returns to schooling

In this section, we adopt the semi-parametric score test to construct confidence bands for the effect of education on wages. To do so, we consider a special case of the LSEM model (3): the linear instrumental variable (IV) model, which has been the workhorse model in the returns to schooling literature (e.g. Card, 2001). We show that the presence of non-Gaussian errors allows us to use the score test to (i) obtain tighter confidence bands for the returns to schooling under the assumption that the instrument is exogenous and (ii) test and correct for possibly endogenous instruments.

We start by showing how the standard linear IV model with control variables can be written as a special case of the general model (3). Let y_i be the dependent variable of interest, w_i the scalar endogenous regressor, z_i the $d_z \times 1$ vectors of instruments and X_i the $d \times 1$ vector of control variables. The linear IV model is given by

$$\begin{aligned} y_i &= \alpha_1 w_i + b_y' X_i + u_i \\ w_i &= \pi' z_i + b_w' X_i + v_i \\ z_i &= B_z X_i + e_i \end{aligned} \quad , \tag{21}$$

where u_i, v_i and e_i are the error terms which are mean zero with variances σ_u^2, σ_v^2 and Σ_e. Further, u_i and v_i are correlated with correlation parameter ρ which captures the endogeneity in the model and prevents us from using basic least squares to estimate α_1. The standard identifying assumption is that e_i is uncorrelated with u_i and v_i such that the instruments given the controls are uncorrelated with the errors.

To write the model in our general notation we first define

$$\begin{bmatrix} u_i \\ v_i \\ e_i \end{bmatrix} = \begin{bmatrix} \sigma_u & 0 & 0 \\ \rho\sigma_v & \sqrt{1-\rho^2}\sigma_v & 0 \\ 0 & 0 & L_e \end{bmatrix} \begin{bmatrix} \epsilon_i^u \\ \epsilon_i^v \\ \epsilon_i^e \end{bmatrix},$$

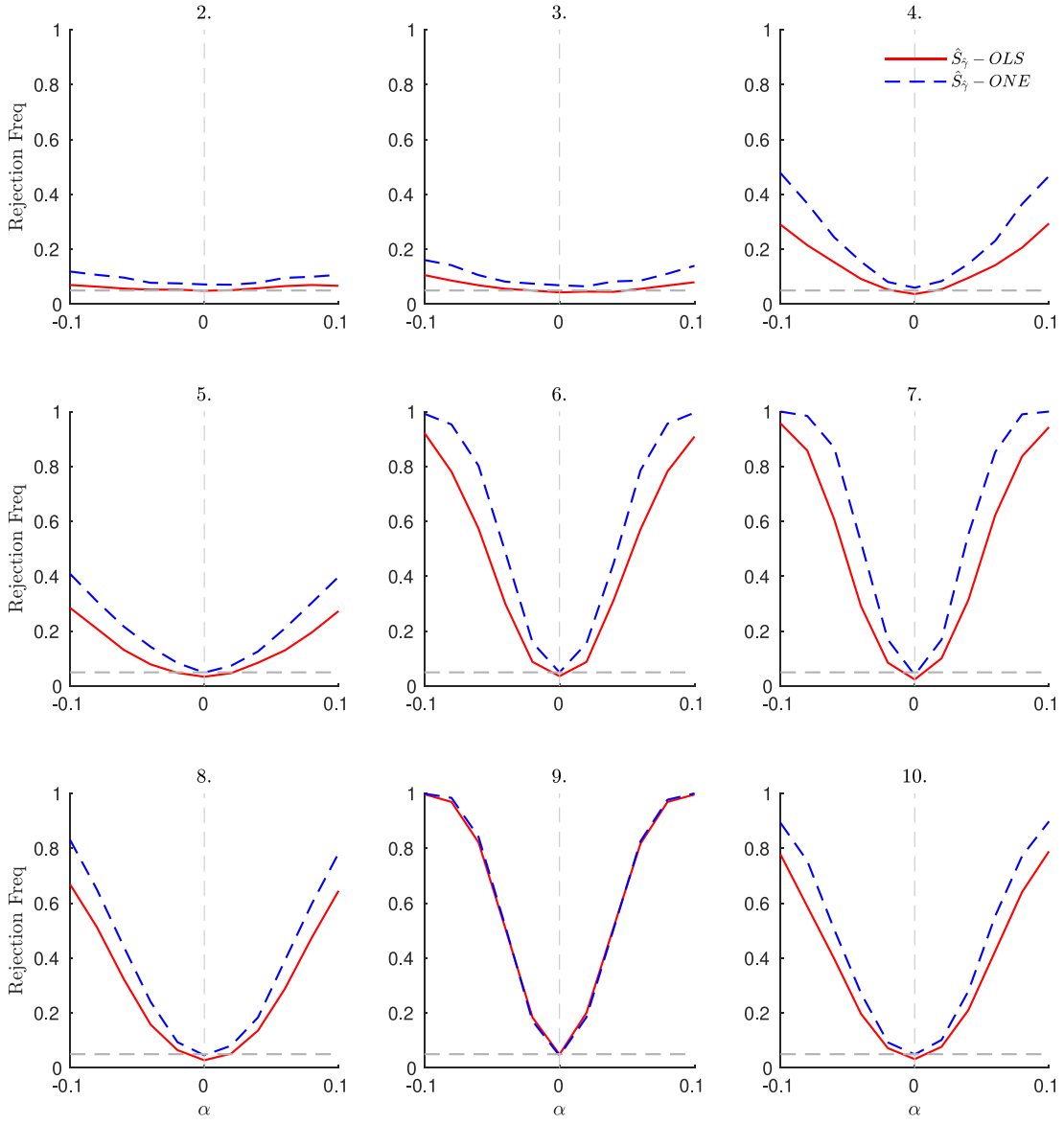

Fig. 5. Power LSEM.

Notes: Empirical power curves for the LSEM model with $k = 2$, $d = 2$ and $n = 1000$. Each plot corresponds to the choice for densities ϵ_{ik}, for $k \geq 2$, where the numbers correspond to the different densities shown in Fig. 3. The solid red line corresponds to the empirical rejection frequency of the $\hat{S}_{\hat{\gamma}}$ test where $\hat{\gamma} = (\alpha_0, \hat{\beta})$, with $\hat{\beta}$ the OLS estimator. The dashed blue line corresponds to the empirical rejection frequency of the $\hat{S}_{\hat{\gamma}}$ test where $\hat{\gamma} = (\alpha_0, \hat{\beta})$, with $\hat{\beta}$ the one-step efficient MLE estimator. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where $\Sigma_e = L_e L_e'$ with L_e lower triangular. To accommodate our general framework we impose that the components of $\epsilon_i = (\epsilon_i^u, \epsilon_i^v, \epsilon_i^e)'$ are mutually independent, with mean zero and unit variance. On this we note that the assumption that the instruments are independent of the error terms u_i and v_i is more commonly imposed (e.g. Hansen et al., 2010; Cattaneo et al., 2012), and below we adopt specification tests to assess whether this assumption is reasonable.

Letting $Y_i = (y_i, w_i, z_i')'$ we have

$$Y_i = BX_i + A^{-1}\epsilon_i\,, \qquad \text{where} \quad A^{-1} = \begin{bmatrix} \sigma_u + \alpha_1\sigma_v\rho & \alpha_1\sqrt{1-\rho^2}\sigma_v & \alpha_1\pi' L_e \\ \rho\sigma_v & \sqrt{1-\rho^2}\sigma_v & \pi' L_e \\ 0 & 0 & L_e \end{bmatrix}, \tag{22}$$

and we set $b = \text{vec}(B)$ and $\sigma = (\pi, \sigma_u, \sigma_v, \rho, \text{vech}(L_e)')'$ to summarize the well identified parameters in our general notation. Model (22) is a special case of the LSEM model (3).

Table 6
Confidence intervals: returns to schooling.

Method	Estimate	Conf interval	Length
$\hat{S}_{\hat{\gamma}}$	–	[0.068 , 0.105]	0.037
AR	–	[0.041 , 0.127]	0.086
OLS	0.076	[0.068 , 0.084]	0.016
2SLS	0.084	[0.040 , 0.127]	0.087

Notes: We report the 95% confidence bands for the effect of education on log wages using the proximity to college interacted with parental education as instrument. The sample size is $n = 2320$ and the model includes control variables for experience, race, smsa and region. The OLS and 2SLS confidence intervals are based on inverting the t-statistic under a normal limiting distribution. The confidence bands corresponding to the semi-parametric score test are based on the $\hat{S}_{\hat{\gamma}}$ implemented using $B = 6$ B-splines and OLS estimates for $\hat{\beta}$. The AR confidence band is based on inverting the Anderson-Rubin statistic.

The parameter α_1 in the linear IV model may not be identified. The standard requirement is that $\pi \neq 0$. However, the current formulation of the linear IV model shows that with non-Gaussian errors we may be able to locally identify α even when the instruments are irrelevant (e.g. Comon, 1994, Theorem 11). More generally, when the instruments are weak but there is a large degree of non-Gaussianity (relative to sampling variation) we may be able to precisely identify α as the instruments are effectively only used to pin down the desired permutation in A.

We emphasize that Theorem 1 ensures that under weak instrument asymptotics, i.e. $\pi = c/\sqrt{n}$ as in Staiger and Stock (1997), the null rejection probability of the semi-parametric score test for testing $H_0 : \alpha = \alpha_0$ does not exceed the nominal level. At the same time we now have two possible identifying sources: the instruments and the non-Gaussian errors. In this sense the model is over-identified and we use this feature below to test the instrument exogeneity condition.

Data. Given this set-up we revisit the returns to schooling problem considered by Card (1995), which uses 1976 wage and schooling data from the 1966 cohort from the NLS to estimate the effect of education on wages. Specifically, for model (22) we set y_i to be the log wage for individual i, w_i is years of eduction, z_i is an indicator for growing up near a 4 year college interacted with parental education and X_i including measures for race, experience, SMSA and region. We refer to Card (2001) for a more general discussion of the literature.

Confidence intervals for the returns to schooling. We start by constructing confidence intervals for α_1 in the model (22) by inverting the semi-parametric score test $\hat{S}_{\hat{\gamma}}$ for the null hypothesis $H_0 : \alpha_1 = \alpha_{1,0}$. We compare this approach to inverting the standard the t-statistic for OLS and 2SLS, as well as inverting the weak instrument robust Anderson-Rubin (AR) statistic. The latter does not exploit non-Gaussian errors but has correct null rejection probability under weak instrument asymptotics (e.g. Staiger and Stock, 1997).

Table 6 shows the different confidence intervals together with the point estimates for OLS and 2SLS. We find that the OLS estimate is smaller when compared to the IV estimate and also has a very small confidence interval resonating with the general findings from Card (2001) that OLS is downward biased and having causal estimates presents a cost in terms of accuracy. The 2SLS and AR confidence bands are very similar as the instrument in this application is strong (the effective F-statistic of Olea and Pflueger (2013) is equal to $F = 80.25$ far exceeding the generalized critical value of 23).

The semi-parametric score test $\hat{S}_{\hat{\gamma}}$ shows the smallest (non – OLS) confidence band for the effect of education on wages $[0.068, 0.105]$, which is considerably smaller when compared to the AR confidence intervals. This reduction in length comes from exploiting non-Gaussian errors in addition to the instrumental variable. Fig. 6 shows kernel density estimates for the residuals from the model, i.e. $\hat{\epsilon}_i = \hat{A}\hat{V}_i$, where $\hat{A} = A(\tilde{\alpha}_1, \hat{\sigma})$ with $\tilde{\alpha}_1$ being the value that minimizes the score statistic. We see that there are modest deviations from the Gaussian distribution which are picked up by the score test and explain the shorter length of the confidence interval.

Instrument validity. A large part of the discussion in Card (1995) and the subsequent literature is devoted to evaluating the validity of the instruments. Several arguments are presented that question the exogeneity of the proximity to schooling instrument. For instance, the presence of a college may be associated with higher school quality in nearby primary and secondary schools, or with geographical variation in wages. Both are not included in the model specification and hence such associations would invalidate the instrument.

To investigate whether the instruments are indeed invalid we extend the model specification for z_i in (21) to allow for correlation with the error term u_i.

$$z_i = B_z X_i + (\alpha_2/\sigma_u)u_i + e_i \ ,$$

where α_2 captures the correlation of the error term with the instrument. The scaling by σ_u is not necessary but makes the LSEM form below slightly more attractive. When $\alpha_2 = 0$ the instrument is exogenous.

With this extension the LSEM parametrization of the IV model becomes

$$Y_i = BX_i + A^{-1}\epsilon_i \ , \quad A^{-1} = \begin{bmatrix} \sigma_u + \alpha_1\sigma_v\rho + \alpha_1\pi'\alpha_2 & \alpha_1\sqrt{1-\rho^2}\sigma_v & \alpha_1\pi' L_e \\ \rho\sigma_v + \pi'\alpha_2 & \sqrt{1-\rho^2}\sigma_v & \pi' L_e \\ \alpha_2 & 0 & L_e \end{bmatrix}, \tag{23}$$

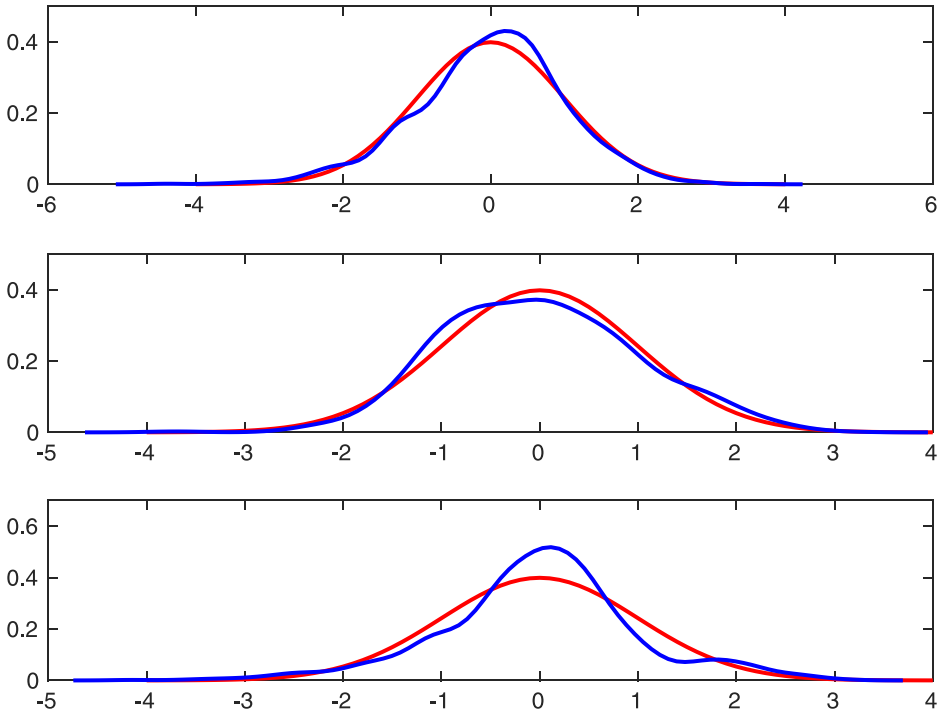

Fig. 6. Densities: returns to schooling.
Notes: We show the kernel density estimates for $\hat{\epsilon}_{i,1}$, $\hat{\epsilon}_{i,2}$ and $\hat{\epsilon}_{i,3}$ (blue line) together with the pdf of the standard normal distribution (red line). The error estimates are obtained as $\hat{\epsilon}_i = \hat{A}\hat{V}_i$, where $\hat{A} = A(\tilde{\alpha}_1, \hat{\sigma})$ with $\tilde{\alpha}_1$ being the value that minimizes the score statistic. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and we test $H_0 : \alpha_1 = \alpha_{1,0}, \alpha_2 = \alpha_{2,0}$ for different values of $\alpha_0 = (\alpha_{1,0}, \alpha_{2,0})$. It is worth pointing out that the inclusion of the additional parameters α_2 prevents the use of standard IV methods, i.e. non-Gaussian errors are needed to distinguish between difference values for α. To do this we use the semi-parametric score test and compare our results to some alternative methods that were discussed in the simulation section.

Fig. 7(a) shows the joint confidence set for α_1 and α_2 that was obtained by inverting $\hat{S}_\gamma$. We find that the hypothesis that the instrument is exogenous (i.e. $\alpha_2 = 0$) cannot be rejected, and the 95% confidence set for α_2 is reasonably tight between approximately -0.2 and 0.25. Most importantly, despite relaxing the instrument validity assumption the implied returns to education are very similar: the confidence set indicates with 95% confidence that the effect of education is between 0.06 and 0.12, only a mild increase when compared to the model that assumes instrument exogeneity.

To showcase the advantage of the semi-parametric score test we also computed a confidence set for α by inverting the pseudo maximum likelihood LM test LM^{pmle} that was discussed in the simulation study, see Fig. 7(b). We find that the confidence set is considerably larger in volume.

Specification tests. We re-emphasize that the semi-parametric score test was build on the underlying assumption that the components of the errors ϵ_i are independent. For the returns to schooling application this implied the errors ϵ_i^u, ϵ_i^v and ϵ_i^z that determine the structural errors and the instruments are independent. To investigate whether this is a plausible assumption we apply the permutation test for mutual independence as proposed by Matteson and Tsay (2017). The p-value for the test is 0.120 and we may conclude that the independence assumption is not rejected for this application, though the evidence is not overwhelming.

In the supplementary material we consider a more general LSEM model which allows for conditional heteroskedasticity. There we repeated the analyses presented here with the difference that the scalings σ_u, σ_v and L_e are allowed to depend on X_i. We find that resulting confidence set for $\alpha = (\alpha_1, \alpha_2)$ is quite similar when compared to its homoskedastic counterpart.

7. Conclusion

In this paper we highlighted a weak identification problem that can arise when non-Gaussianity is used to identify parameters in LSEMs. The consequence of this problem is that several existing inference methods suffer from size distortions when the true distributions are close to Gaussian.

To reduce this problem we proposed a semi-parametric score statistic for testing hypotheses in LSEMs. Under mild regularity conditions we demonstrated that the semi-parametric score test is locally robust in the sense that its null rejection probability is

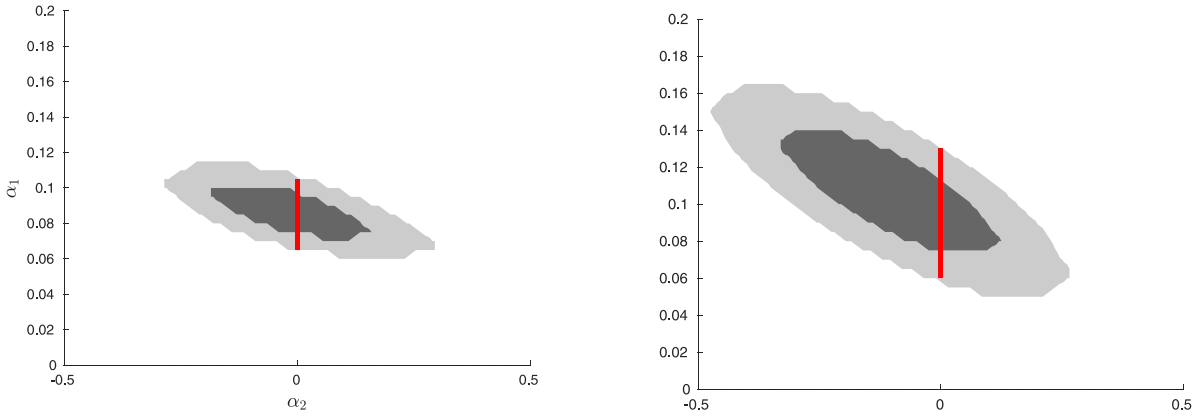

Fig. 7. Confidence sets: returns to schooling.
Notes: We show 95% (light gray) and 67% (dark gray) confidence sets for $\alpha = (\alpha_1, \alpha_2)$, where α_1 captures the effect of education on log wages and α_2 captures the correlation between the instrument (proximity to schooling interacted with parental education) and the error of the log wage equation. The red line indicates the confidence interval under the restriction of instrument exogeneity, i.e. $\alpha_2 = 0$. Figure (a) shows the result after inverting the weak non-Gaussianity robust test $\hat{S}_{\hat{\gamma}}$. Figure (b) shows the result after inverting the pseudo MLE LM test based on the Student's t density. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

no greater than the nominal level under parameter sequences that can be described by local deviations from the true parameters which satisfy the null hypothesis (i.e. under weak identification asymptotics). A simulation study shows that our asymptotic theory provides an accurate approximation to the finite sample performance of our test.

While we have restricted our treatment to models where the observations were independently distributed across entities, we note that a similar approach may be considered for dynamic models, but this will require extending our results to allow for non-i.i.d. data. Further, whilst our work shows that the semi-parametric score test is robust under weak identification asymptotics, no global uniformity results are derived. These extensions are left for future work.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2023.105647.

References

Amari, S.-I., Cardoso, J.-F., 1997. Blind source separation-semiparametric statistical approach. IEEE Trans. Signal Process. 45 (11), 2692–2700.
Anderson, T.W., Rubin, H., 1949. Estimation of the parameters of a single equation in a complete system of stochastic equations. Ann. Math. Stat. 20 (1), 46–63.
Andrews, D.W.K., 1987. Asymptotic results for generalized wald tests. Econom. Theory 3 (3), 348–358.
Andrews, D.W.K., Cheng, X., 2012. Estimation and inference with weak, semi-strong, and strong identification. Econometrica 80 (5), 2153–2211.
Andrews, D.W.K., Cheng, X., 2013. Maximum likelihood estimation and uniform inference with sporadic identification failure. J. Econometrics 173 (1), 36–56.
Andrews, D.W.K., Cheng, X., Guggenberger, P., 2020. Generic results for establishing the asymptotic size of confidence sets and tests. J. Econometrics 218 (2), 496–531.
Andrews, D.W.K., Guggenberger, P., 2019. Identification- and singularity-robust inference for moment condition models. Quant. Econ. 10 (4), 1703–1746.
Andrews, I., Mikusheva, A., 2015. Maximum likelihood inference in weakly identified dynamic stochastic general equilibrium models. Quant. Econ. 6 (1), 123–152.
Andrews, I., Mikusheva, A., 2022. Optimal decision rules for weak GMM. Econometrica 90 (2), 715–748.
Azzalini, A., Capitanio, A., 2014. The Skew-Normal and Related Families. Cambridge University Press.
Bekaert, G., Engstrom, E., Ermolov, A., 2020. Aggregate Demand and Aggregate Supply Effects of COVID-19: A Real-time Analysis. Working paper.
Bekaert, G., Engstrom, E., Ermolov, A., 2021. Macro risks and the term structure of interest rates. J. Financ. Econ. 141 (2), 479–504.
Bickel, P.J., Klaasen, C.A.J., Ritov, Y., Wellner, J.A., 1998. Efficient and Adaptive Estimation for Semiparametric Models. Springer, New York, NY, USA.
Bickel, P.J., Ritov, Y., Stoker, T.M., 2006. Tailor-made tests for goodness of fit to semiparametric hypotheses. Ann. Statist. 34 (2), 721–741.
Bonhomme, S., Robin, J.-M., 2009. Consistent noisy independent component analysis. J. Econometrics 149 (1), 12–25.
Card, D., 1995. Using geographic variation in college proximity to estimate the return to schooling. In: Christofides, Louis N., Grant, E. Kenneth, Swidinsky, Robert (Eds.), Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp. University of Toronto Press, pp. 201–222.
Card, D., 2001. Estimating the return to schooling: Progress on some persistent econometric problems. Econometrica 69 (5), 1127–1160.
Cattaneo, M.D., Crump, R.K., Jansson, M., 2012. Optimal inference for instrumental variables regression with non-Gaussian errors. J. Econometrics 167 (1), 1–15.
Chen, A., Bickel, P.J., 2006. Efficient independent component analysis. Ann. Statist. 34 (6), 2825–2855.
Choi, S., Hall, W.J., Schick, A., 1996. Asymptotically uniformly most powerful tests in parametric and semiparametric models. Ann. Statist. 24 (2), 841–861.
Comon, P., 1994. Independent component analysis, a new concept? Signal Process. 36 (3), 287–314.
Dagenais, M.G., Dagenais, D.L., 1997. Higher moment estimators for linear regression models with errors in the variables. J. Econometrics 76 (1–2), 193–221.
Davis, R., Fernandes, L., 2022. Independent Component Analysis with Heavy Tails using Distance Covariance. Working paper.
Davis, R., Ng, S., 2022. Time series estimation of the dynamic effects of disaster-type shocks. J. Econometrics 235 (1), 180–201.
de Boor, C., 2001. A Practical Guide to Splines. Springer, New York, NY, USA.
Dhrymes, P.J., 1994. Topics in Advanced Econometrics, Volume II Linear and Nonlinear Simultaneous Equations. Springer-Verlag New York.

Drautzburg, T., Wright, J.H., 2023. Refining set-identification in VARs through independence. J. Econometrics 235 (2), 1827–1847.
Erickson, T., Whited, T.M., 2000. Measurement error and the relationship between investment and q. J. Polit. Econ. 108 (5), 1027–1057.
Erickson, T., Whited, T.M., 2002. Two-step GMM estimation of the errors-in-variable model using higher order moments. Econom. Theory 18 (3), 776–799.
Frisch, R., 1933. Propagation problems and impulse problems in dynamic economics. In: Economic Essays in Honor of Gustav Cassel. George Allen and Unwin.
Gabriele, F., Sentana, E., 2023. Discrete mixtures of normals pseudo maximum likelihood estimators of structural vector autoregressions. J. Econometrics 235 (2), 643–665.
Gouriéroux, C., Monfort, A., Renne, J.-P., 2017. Statistical inference for independent component analysis: Application to structural VAR models. J. Econometrics 196, 111–126.
Gouriéroux, C., Monfort, A., Renne, J.-P., 2019. Identification and estimation in non-fundamental structural VARMA models. Rev. Econom. Stud. 87 (4), 1915–1953.
Granziera, E., Moon, H.R., Schorfheide, F., 2018. Inference for VARs identified with sign restrictions. Quant. Econ. 9 (3), 1087–1121.
Guay, A., 2021. Identification of structural vector autoregressions through higher unconditional moments. J. Econometrics 225 (1), 27–46.
Haavelmo, T., 1943. The statistical implications of a system of simultaneous equations. Econometrica 11, 1–12.
Haavelmo, T., 1944. The probability approach in econometrics. Econometrica 12, Supplement.
Hall, W.J., Mathiason, D.J., 1990. On large-sample estimation and testing in parametric models. Internat. Statist. Rev. 58 (1), 77–97.
Hansen, C., McDonald, J.B., Newey, W.K., 2010. Instrumental variables estimation with flexible distributions. J. Bus. Econom. Statist. 28 (1), 13–25.
Herwartz, H., 2019. Long-run neutrality of demand shocks: Revisiting Blanchard and Quah (1989) with independent structural shocks. J. Appl. Econometrics 34 (5), 811–819.
Herwartz, H., Lange, A., Maxand, S., 2019. Statistical Identification in Svars - Monte Carlo Experiments and a Comparative Assessment of the Role of Economic Uncertainties for the US Business Cycle. CEGE Discussion Paper 375, CEGE.
Horn, R.A., Johnson, C.R., 2013. Matrix Analysis, second ed. Cambridge University Press.
Hyvärinen, A., Karhunen, J., Oja, E., 2001. Independent Component Analysis. John Wiley & Sons, Inc..
Jin, K., 1992. Empirical smoothing parameter selection in adaptive estimation. Ann. Statist. 20 (4), 1844–1874.
Jin, Z., Risk, B.B., Matteson, D.S., 2019. Optimization and testing in linear non-Gaussian component analysis. Stat. Anal. Data Min. ASA Data Sci. J. 12 (3), 141–156.
Kaji, T., 2021. Theory of weak identification in semiparametric models. Econometrica 89 (2), 733–763.
Kapteyn, A., Wansbeek, T., 1983. Identification in the linear errors in variables model. Econometrica 51 (6), 1847–1849.
Kleibergen, F., 2005. Testing parameters in GMM without assuming that they are identified. Econometrica 73 (4), 1103–1123.
Kocherlakota, S., Kocherlakota, K., 1991. Neyman's C(α) test and Rao's efficient score test for composite hypotheses. Statist. Probab. Lett. 11 (6), 491–493.
Lanne, M., Luoto, J., 2021. GMM estimation of non-Gaussian structural vector autoregression. J. Bus. Econom. Statist. 39 (1), 69–81.
Lanne, M., Lütkepohl, H., 2010. Structural vector autoregressions with nonnormal residuals. J. Bus. Econom. Statist. 28 (1), 159–168.
Lanne, M., Meitz, M., Saikkonen, P., 2017. Identification and estimation of non-Gaussian structual vector autoregressions. J. Econometrics 196, 288–304.
Le Cam, L.M., 1960. Locally Asymptotically Normal Families of Distributions: Certain Approximations to Families of Distributions and their Use in the Theory of Estimation and Testing Hypotheses. University of California Berkeley, Calif: University of California publications in statistics, University of California Press.
Le Cam, L.M., Yang, G.L., 2000. Asypmtotics in Statistics: Some Basic Concepts, second ed. Springer, New York, NY, USA.
Lee, A., 2023. Robust and Efficient Inference for Non-Regular Semiparametric Models. Working paper.
Lehmann, E.L., Romano, J.P., 2005. Testing Statistical Hypotheses, third ed. Springer, New York, NY, USA.
Lewbel, A., Schennach, S.M., Zhang, L., 2023. Identification of a triangular two equation system without instruments. J. Bus. Econom. Statist. (forthcoming).
Lütkepohl, H., Burda, M.M., 1997. Modified wald tests under nonregular conditions. J. Econometrics 78 (2), 315–332.
Magnus, J.R., Pijls, H.G.J., Sentana, E., 2021. The Jacobian of the exponential function. J. Econom. Dynam. Control 127, 104122.
Marron, J.S., Wand, M.P., 1992. Exact mean integrated squared error. Ann. Statist. 20 (2), 712–736.
Matteson, D.S., Tsay, R.S., 2017. Independent component analysis via distance covariance. J. Amer. Statist. Assoc. 112 (518), 623–637.
Maxand, S., 2018. Identification of independent structural shocks in the presence of multiple Gaussian components. Econ. Statist. 55–68.
Moneta, A., Entner, D., Hoyer, P.O., Coad, A., 2013. Causal inference by independent component analysis: Theory and applications*. Oxf. Bull. Econ. Statist. 75 (5), 705–730.
Moneta, A., Pallante, G., 2022. Identification of structural VAR models via independent component analysis: A performance evaluation study. J. Econom. Dynam. Control 144, 104530.
Moreira, M.J., 2003. A conditional likelihood ratio test for structual models. Econometrica 71 (4), 1027–1048.
Newey, W.K., 1990. Semiparametric efficiency bounds. J. Appl. Econometrics 5 (2), 99–135.
Neyman, J., 1979. C(α) tests and their use. Sankhyā 41 (1/2), 1–21.
Olea, J.L. Montiel, Pflueger, C., 2013. A robust test for weak instruments. J. Bus. Econom. Statist. 31 (3), 358–369.
Reiersøl, O., 1950. Identifiability of a linear relation between variables which are subject to error. Econometrica 18 (4), 375–389.
Risk, B.B., Matteson, D.S., Ruppert, D., 2019. Linear non-Gaussian component analysis via maximum likelihood. J. Amer. Statist. Assoc. 114 (525), 332–343.
Rothenberg, T.J., 1971. Identification in parametric models. Econometrica 39 (3), 577–591.
Sen, A., 2012. On the Interrelation Between the Sample Mean and the Sample Variance. Amer. Statist. 66 (2), 112–117.
Sims, C.A., 2021. SVAR Identification through Heteroskedasticity with Misspecified Regimes. Working Paper.
Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. Econometrica 65 (3), 557–586.
Stock, J.H., Wright, J.H., 2000. GMM with weak identification. Econometrica 68 (5), 1055–1096.
Tank, A., Fox, E.B., Shojaie, A., 2019. Identifiability and estimation of structural vector autoregressive models for subsampled and mixed-frequency time series. Biometrika 106 (2), 433–452.
Tinbergen, J., 1939. Statistical Testing of Business Cycle Theories: Part I: A Method and Its Application to Investment Activity.
van der Vaart, A.W., 1998. Asymptotic Statistics, first ed. Cambridge University Press, New York, NY, USA.
van der Vaart, A.W., 2002. Semiparametric statistics. In: Bernard, P. (Ed.), Lectures on Probability Theory and Statistics: Ecole d'Eté de Probabilités de Saint-Flour XXIX - 1999. Springer, Berlin, Germany.
Velasco, C., 2023. Identification and estimation of structural VARMA models using higher order dynamics. J. Bus. Econom. Statist. 41 (3), 819–832.
Working, E.J., 1927. What do statistical "demand curves" show? Q. J. Econ. 41 (2), 212–235.