A Geometric Approach to Nonlinear Econometric Models

A Geometric Approach to Nonlinear Econometric Models

By Isaiah Andrews[1] and Anna Mikusheva[2]

Abstract

Conventional tests for composite hypotheses in minimum distance models can be unreliable when the relationship between the structural and reduced-form parameters is highly nonlinear. Such nonlinearity may arise for a variety of reasons, including weak identification. In this note we begin by studying the problem of testing a "curved null" in a finite-sample Gaussian model. Using the curvature of the model we develop new finite-sample bounds on the distribution of minimum-distance statistics. These bounds allow us to construct tests for composite hypotheses which are uniformly asymptotically valid over a large class of data generating processes and structural models.

Key words: weak identification, statistical differential geometry

First draft: May 2012

This draft: November, 2015.

1 Introduction

Economists frequently fit nonlinear models using minimum-distance techniques, which attempt to match model predictions to reduced-form parameter estimates. Conventional tests for composite hypotheses in this setting implicitly rely on linear approximations, such as the delta-method and first-order Taylor expansions, and can be unreliable in contexts where nonlinearity is important relative to the variability of the reduced-form parameter estimates. Such nonlinearity may arise for a variety of reasons, for example if the model being fitted, or the restriction under test, is highly nonlinear relative to the sample size. Relatedly, nonlinearity may result from weak identification of structural parameters. In this paper we develop techniques for inference which are robust to nonlinearity in the relationship between the structural and reduced-form parameters.

[1]Harvard Society of Fellows. Harvard Department of Economics, Littauer Center 124, Cambridge, MA 02138. Email iandrews@fas.harvard.edu. NSF Graduate Research Fellowship support under grant number 1122374 is gratefully acknowledged.

[2]Department of Economics, M.I.T., 50 Memorial Drive, Building E52, Cambridge, MA, 02142. Email: amikushe@mit.edu. Financial support from the Castle-Krob Career Development Chair and Sloan Research Fellowship is gratefully acknowledged. We are extremely grateful to Professor of Differential Geometry Tobias Colding for discussing geometric issues with us, and for checking our geometric proofs. We are grateful to Gary Chamberlain, Victor Chernozhukov, Jerry Hausman, Mattias Cattaneo, Whitney Newey, Jim Stock, co-editor and several anonymous referees for helpful comments. A previous version of this paper circulated under the title "A Geometric Approach to Weakly Identified Econometric Models."

We first study the problem of testing a nonlinear hypothesis on the mean of a Gaussian vector with unknown mean and known variance. We show that the distribution of minimum-distance statistics in this context is dominated by an easy-to-simulate distribution which depends only the geometric curvature of a manifold defined by the null hypothesis, measured relative to the known variance matrix. Using this bound, we derive a novel test which controls finite-sample size. This test is always more powerful than tests based on the projection method, a leading alternative for this problem. Our approach differs from the statistical geometry literature initiated by Efron (1975) in that we produce finite sample bounds on the distribution of the test statistic, whereas the statistical geometry literature is primarily concerned with higher-order asymptotic approximations.

We show that our finite-sample bounds allow us to derive uniformly asymptotically valid minimum-distance tests. These tests control size uniformly over a large class of data-generating processes and structural models, allowing arbitrarily nonlinear relationships between the reduced-form and structural parameters. Moreover, in cases where conventional linear asymptotic approximations are reliable our robust tests coincide with conventional tests asymptotically and thus do not sacrifice asymptotic power in these cases. We also introduce two modifications of our baseline procedure which offer computational and power advantages in many contexts. Implementing our tests requires only solving a non-stochastic optimization problem to compute geometric curvature and so does not entail repeated simulation of the minimum distance statistic.

The paper is structured as follows. In Section 2 we introduce a finite-sample testing problem, derive geometric and statistical bounds, and introduce our baseline test. In Section 3, we extend our finite-sample results to show uniform asymptotic validity, discuss the behavior of our test under conventional asymptotics, and compare our approach to existing alternatives. Section 4 introduces two modifications of our baseline procedure and discusses implementation. All proofs may be found in the Supplementary Appendix, available on Anna Mikusheva's website.[3] In the Supplement we further show analytically that weak identification leads to asymptotic nonlinearity in a toy DSGE example.

We use the following notation: $\dot{\gamma}$ is the derivative of the function γ, $\ddot{\gamma}$ is the second derivative, $B_R(x_0) = \{x \in \mathbb{R}^k : \|x - x_0\| \leq (1+\sqrt{2})R\}$ is a k-dimensional ball of radius $(1+\sqrt{2})R$ with center x_0, and $|A|$ is the cardinality of a set A.

[3] `economics.mit.edu/faculty/amikushe`

2 Finite-sample Inference in a Gaussian Model

We consider the problem of testing a potentially non-linear hypothesis on the mean of a multivariate Gaussian vector. Assume we observe a k-dimensional Gaussian vector $\widehat{\theta}$ with known covariance matrix Σ and unknown mean θ_0. We wish to test a p-dimensional restriction which may be formulated either as $g(\theta_0) = 0$ for some $(k-p)$-dimensional smooth function g or as $\theta_0 = \theta(\beta_0)$ for a known link function $\theta(\cdot)$ and some unknown p-dimensional parameter β_0 lying in a parameter space $U \subseteq \mathbb{R}^p$.

Such testing problems arise in many contexts, for example in testing hypotheses with nuisance parameters or in testing model specification. This is the limiting testing problem in many weakly identified minimum distance models, as well as cases when one fits a highly nonlinear structural model based on reduced-form parameter estimates. For example, let $\widehat{\theta}$ be a preliminary or reduced-form estimator, which is approximately normal with a well-estimable covariance matrix Σ. Assume the relationship between the structural and reduced-form parameters is described by the link function $\theta(\beta)$ for structural parameter β. Then testing correct model specification is (asymptotically) equivalent to testing that $\theta_0 = \theta(\beta_0)$ for some $\beta_0 \in U$. Alternatively, if there are two structural parameters λ and β with link function $\theta(\lambda, \beta)$, then testing a hypothesis about λ alone, $H_0 : \lambda = \lambda_0$, is equivalent to testing $H_0 : \theta_0 = \theta(\lambda_0, \beta)$ for some $\beta \in U$. We base inference on the minimum distance (or for the exact Gaussian case, likelihood ratio) statistic, which may be formulated as

$$MD = \min_{\theta: g(\theta)=0} (\widehat{\theta} - \theta)' \Sigma^{-1} (\widehat{\theta} - \theta_0) \text{ or } MD = \min_{\beta \in U} (\widehat{\theta} - \theta(\beta))' \Sigma^{-1} (\widehat{\theta} - \theta(\beta))$$

depending on the formulation of the null hypothesis.

To proceed, let us introduce the normalized random vector $\xi = \Sigma^{-1/2}(\widehat{\theta} - \theta_0) \sim N(0, I_k)$ and the p-dimensional manifold $S = \{x : x = \Sigma^{-1/2}(\theta(\beta) - \theta_0), \beta \in \mathbb{R}^p\}$ or $S = \{x : x = \Sigma^{-1/2}(\theta - \theta_0), g(\theta) = 0\}$. Note that the manifold S is known up to a location shift determined by the true value θ_0. Thus, we know the shape of S and, moreover, know that it passes through the origin if the null holds. The minimum distance statistics defined above are simply the squared distance between ξ and S:

$$MD = \min_{x \in S} (\xi - x)'(\xi - x) = \rho^2(\xi, S), \tag{1}$$

where ρ is the Euclidean distance from a point to a set. The distribution of $\rho^2(\xi, S)$ is in general non-standard and depends on the unknown θ_0.

The central issue of this paper is how to find computationally tractable critical values

such that tests based on $\rho^2(\xi, S)$ control size. The problem is that we do not know the location of the true value θ_0 on the null manifold, but the distribution of the statistic depends on this value. Thus, we face a nuisance parameter problem.

We will distinguish between the linear and non-linear cases. If S is a p-dimensional linear sub-space in $\mathbb{R}^k$ then the squared distance $\rho^2(\xi, S)$ has a χ^2_{k-p} distribution. Most of the classical statistics literature deals with testing hypotheses that are either linear or asymptotically linear, in the sense that S is either a linear subspace or is well-approximated by one in large samples. In particular, classical delta-method arguments assume that the reduced-form parameter is precisely estimated relative to the nonlinearity of the null hypothesis manifold, and thus that we can linearize the null hypothesis manifold around the true parameter value.

By contrast, we also want to allow cases where the non-linearity of the model is important relative to the sampling error of the reduced-form parameter estimates, rendering linear approximations unreliable. As noted by Hansen and Sargent (1991) in a discussion of rational expectations models, "even for models that are linear in the variables the cross-equation restrictions on the parameters can be complicated and often highly non-linear". Another potential source of nonlinearity in S is weak identification: in the Supplementary Appendix we study an analytic DSGE example in which weak identification, arising from insufficiently rich time-series dynamics for structural shocks, means that hypotheses about structural parameters yield highly nonlinear null hypothesis manifolds.

One bound can be placed on $\rho^2(\xi, S)$ without any assumptions, namely that $\rho^2(\xi, S)$ is dominated by a χ^2_k distribution. Indeed, since $0 \in S$,

$$\rho(\xi, S)^2 = \min_{x \in S} (\xi - x)'(\xi - x) \leq (\xi - 0)'(\xi - 0) \sim \chi^2_k. \tag{2}$$

Using this bound gives the "projection method," which is currently the main approach available for testing with weakly identified nuisance parameters, see Dufour and Jasiak (2001), and Dufour and Taamouti (2005). This paper proposes new critical values based on a stochastic bound on the distribution of the MD statistic. These critical values are smaller than those used by the projection method and coincide with χ^2_{k-p} critical values for linear hypotheses. This bound is based on measuring the curvature of the null hypothesis relative to the variance Σ of the reduced-form parameter estimates.

2.1 Geometric Concepts

In this paper we focus on regular manifolds embedded in k-dimensional Euclidean space. A subset $S \subset \mathbb{R}^k$ is called a *p-dimensional regular manifold* if for each point $q \in S$ there exists a neighborhood V in $\mathbb{R}^k$ and a twice-continuously-differentiable map $\mathbf{x} : \widetilde{U} \to V \bigcap S$ from an open set $\widetilde{U} \subset \mathbb{R}^p$ onto $V \bigcap S \subset \mathbb{R}^k$ such that (i) $\mathbf{x}$ is a homeomorphism, which is to say it has a continuous inverse and (ii) the Jacobian $d\mathbf{x}_q$ has full rank. A mapping $\mathbf{x}$ which satisfies these conditions is called a parametrization or a system of local coordinates.

For $\mathbf{x}$ a system of local coordinates at q, the set of all tangent vectors to S at q coincides with the linear space spanned by the Jacobian $d\mathbf{x}_q$ and is called the *tangent space* to S at q (denoted $T_q(S)$). Denote by $\gamma : (-\varepsilon, \varepsilon) \to S$ a curve which lies in S and passes through $q = \gamma(0)$. The measure of curvature we consider is

$$\kappa_q(S) = \sup_{X \in T_q(S), \dot{\gamma}(0) = X} \kappa_q(\gamma, S) = \sup_{X \in T_q(S), \dot{\gamma}(0) = X} \frac{\left\| (\ddot{\gamma}(0))^{\perp} \right\|}{\|\dot{\gamma}(0)\|^2}, \tag{3}$$

where $(W)^{\perp}$ stands for the projection of W onto the space orthogonal to $T_q(S)$. This measure of curvature is equal to the maximal curvature over all geodesics passing through the point q and is invariant to the parametrization. If S is a p-dimensional sphere of radius C then for each $q \in S$ we have $\kappa_q(S) = 1/C$. If, on the other hand, S is a linear subspace its curvature is zero at all points. Further discussion of geometric concepts is deferred to the Supplementary Appendix.

How to calculate curvature in practice. Let S be a p-dimensional manifold in $\mathbb{R}^k$, and $\mathbf{x}$ a local parametrization at a point q, $q = \mathbf{x}(y^*)$. Denote the derivatives of $\mathbf{x}$ at q by $v_i = \frac{\partial \mathbf{x}}{\partial y_i}(y^*)$, and let $Z = (v_1, ..., v_p)$. For any vector $W \in \mathbb{R}^k$ let $W^{\perp} = N_Z W = (I - Z(Z'Z)^{-1}Z')W$. Finally, denote the p^2 vectors of second derivatives $V_{ij} = \frac{\partial^2}{\partial y_i \partial y_j}\mathbf{x}(y^*)$. The curvature can then be written as

$$\kappa_q(S) = \sup_{\substack{u=(u_1,..,u_p)\in\mathbb{R}^p \\ \|\sum_{i=1}^p u_i v_i\|=1}} \left\| \sum_{i,j=1}^{p} u_i u_j V_{ij}^{\perp} \right\| = \sup_{(w_1,...,w_p)\in\mathbb{R}^p} \frac{\left\| \sum_{i,j=1}^p w_i w_j V_{ij}^{\perp} \right\|}{\left\| \sum_{i=1}^p w_i v_i \right\|^2}. \tag{4}$$

2.2 Geometric Bounds

We bound the distance in $\mathbb{R}^k$ from a random vector $\xi \sim N(0, I_k)$ to a p-dimensional non-random manifold S that contains zero. Our bound depends on the maximal curvature

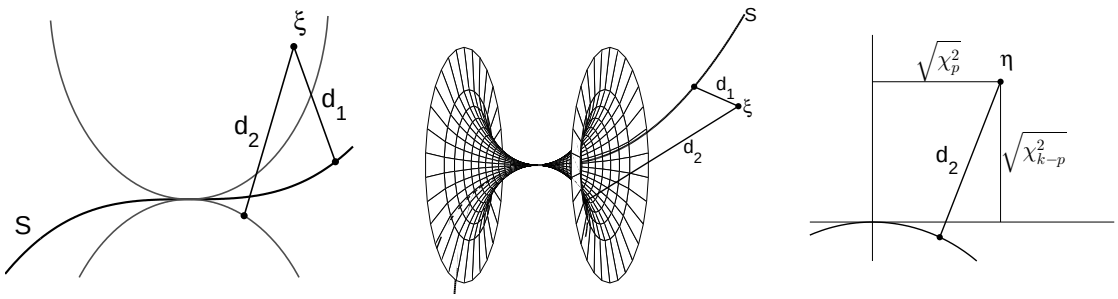

Figure 1: Left panel: Bounding a line between two curves in Case 1. Middle panel: the envelope for a space curve in $\mathbb{R}^3$ in Case 2. Right panel: The stochastic bound described in Theorem 1 (d).

$\kappa_q(S)$ over all relevant points in the manifold S. The bound depends on global properties of the manifold, in the sense of properties that hold on a fixed bounded set, but the behavior of the manifold at infinity is irrelevant. In what follows, we restrict attention to a connected part of the manifold that lies inside of a finite cylinder centered at zero.

We derive our bound in two steps: first, we construct an envelope for the manifold S using a collection of p-dimensional spheres. We show that the distance from any point ξ to S is bounded above by the distance from ξ to the most distant sphere in the collection we consider. Second, we show that our geometric construction implies a bound on the distribution of $\rho^2(\xi, S)$ and hence on the distribution of the minimum distance statistic. To provide intuition for our main statement we first discuss two simple cases in which the construction of the envelope can be easily visualized.

Case 1 (k=2, p=1): A curve in $\mathbb{R}^2$. Consider a curve S passing through zero (i.e. $(0,0) \in S$). Suppose that the curvature of S is less than or equal to $1/C$ for all points in S. If we imagine two circles of radius C tangent to S at zero, we can see that S lies between them- see the left panel of Figure 1 for illustration. The distance from any point ξ to S (denoted by d_1 in the left panel of Figure 1) does not exceed the distance from ξ to the further of the two circles (denoted by d_2). This is the geometrical bound we use.

Note that if the maximal curvature of S goes to zero at all points (so that $C \to \infty$) then the two bounding circles converge to the tangent line to S at zero on any bounded set. Further, note that the distribution of the distance d_2 from a normal random vector to the furthest of two circles depends only on C and is easy to simulate.

The logic of this example is quite straightforward to generalize to the case of a $k-1$-dimensional manifold in $\mathbb{R}^k$, known as a hyper-surface. Dealing with manifolds of lower dimension is more challenging, but the basic principle of the approach can be illustrated

using a curve in $\mathbb{R}^3$.

Case 2 (k=3, p=1): A curve in $\mathbb{R}^3$. Suppose now that we have a one-dimensional space curve S in $\mathbb{R}^3$ that passes through zero and whose curvature at all points is bounded above by $1/C$. We construct our envelope by considering the collection of all one-dimensional circles of radius C tangent to S at zero. Equivalently, one can take a given circle tangent to S at zero and rotate it around the tangent line. An example of the resulting surface is given on the middle panel of Figure 1: we can again see that the curve S lies inside the envelope.

One can show that the distance from any point ξ to the curve S (denoted by d_1 in the middle panel of Figure 1) is bounded above by the distance from ξ to the furthest circle in the collection used to construct the envelope (denoted d_2). Note that if the curvature of S goes to zero at all points (so that $C \to \infty$) then on any bounded set the envelope converges to the tangent line to S at zero.

This geometric bound immediately implies a bound on the distribution of $\rho^2(\xi, S)$. For $\xi \sim N(0, I_3)$ the distribution of the distance d_2 from ξ to the furthest circle is simple to simulate. One can show that it is distributed as the squared distance from a two-dimensional random vector η to the circle of radius C with center $(0, -C)$ where the coordinates of η are distributed as independent $\sqrt{\chi_1^2}$ and $\sqrt{\chi_2^2}$ random variables.

General case With the intuition provided by these examples, we now turn to the general case. Let S be a regular connected p-dimensional manifold in $\mathbb{R}^k$ passing through zero. By the rotation invariance of standard normal vectors we can assume without loss of generality that the tangent space $T_0(S)$ to manifold S at zero is spanned by the first p basis vectors. For each $x \in \mathbb{R}^k$, let $x = (x^{(1)}, x^{(2)})$ where $x^{(1)} = (x_1, ..., x_p) \in \mathbb{R}^p$ contains the first p coordinates of x while $x^{(2)} = (x_{p+1}, ..., x_k) \in \mathbb{R}^{k-p}$ contains the last $k-p$. In what follows, we restrict attention to points on the manifold that lie inside of a (large) finite cylinder $D_C = \{x = (x^{(1)}, x^{(2)}) : \|x^{(1)}\| \leq C, \|x^{(2)}\| \leq C, x^{(1)} \in \mathbb{R}^p, x^{(2)} \in \mathbb{R}^{k-p}\} \subset \mathbb{R}^k$. Let S_C be the intersection $S \bigcap D_C$ if it is connected or the *connected* part of $S \bigcap D_C$ that passes through zero (that is, the part of $S \bigcap D_C$ which can be reached by continuous paths lying in $S \bigcap D_C$ which pass through zero) otherwise. Note that $\rho(\xi, S) \leq \rho(\xi, S_C)$.

To obtain some of our bounding results, we need one further assumption:

Assumption 1 *For any $y^{(1)} \in \mathbb{R}^p$ with $\left\|y^{(1)}\right\| \leq C$ there exists a point $x \in S_C$ such that $x^{(1)} = y^{(1)}$.*

Since we assumed (without loss of generality) that the tangent space $T_0(S)$ is spanned by the first p basis vectors, Assumption 1 requires that the projection of S_C on its tangent space at zero cover a p-dimensional ball of radius C centered at zero, and hence that S_C have dimension p in a global sense. By a local property we mean one that holds on an infinitesimal neighborhood of a point. In contrast, by a global property we mean one that holds on a fixed bounded set. Lemma 1 shows that Assumption 1 holds quite generally for implicitly defined manifolds.

Lemma 1 *Let the p-dimensional manifold S in $\mathbb{R}^k$ be defined by $S = \left\{x \in \mathbb{R}^k, g(x) = 0\right\}$ for a continuously differentiable function $g : \mathbb{R}^k \to \mathbb{R}^{k-p}$. Assume that $g(\mathbf{0}_k) = 0$. For some $C > 0$ assume that $\frac{\partial}{\partial x'} g(x)$ is full rank for all $x \in S_C$. If the maximal curvature over S_C is not larger than $1/C$, then the projection of S_C on the tangent space $T_0(S_C)$ covers the ball of radius C centered at zero.*

Theorem 1 *Let S be a regular p-dimensional manifold in $\mathbb{R}^k$ passing through zero. Assume that the tangent space $T_0(S)$ is spanned by the first p basis vectors. Assume that for some constant $C > 0$ we have that $\kappa_q(S) \leq \frac{1}{C}$ for all points $q \in S_C$. Then:*

(a) Manifold S_C lies inside the set $\mathcal{M} \cap D_C$, where

$$\mathcal{M} = \{\|x^{(1)}\|^2 + (C - \|x^{(2)}\|)^2 \geq C^2\}. \tag{5}$$

(b) If Assumption 1 is satisfied, then for any point $\xi \in \mathbb{R}^k$ we have almost surely

$$\rho(\xi, S) \leq \max_{u \in \mathbb{R}^{p-k}, \|u\|=1} \rho(\xi, N_u),$$

where $N_u = \{x \in \mathbb{R}^k : x = (x^{(1)}, zu), x^{(1)} \in \mathbb{R}^p, z \in \mathbb{R}_+, \|x^{(1)}\|^2 + (C - z)^2 = C^2\}$.

(c) Almost surely $\max_{u \in \mathbb{R}^{p-k}, \|u\|=1} \rho(\xi, N_u) = \rho(\xi, N_{\widetilde{u}})$, where $\widetilde{u} = -\frac{1}{\|\xi^{(2)}\|}\xi^{(2)}$.

(d) If $\xi \sim N(0, I_k)$ we have for all x, y:

$$P\left\{\max_{u \in \mathbb{R}^{p-k}, \|u\|=1} \rho^2(\xi, N_u) \leq x, \|\xi\| \leq y\right\} = P\left\{\rho_2^2(\eta, N_2^C) \leq x, \|\eta\| \leq y\right\},$$

where the coordinates of the 2-dimensional random vector $\eta = (\sqrt{\chi_p^2}, \sqrt{\chi_{k-p}^2}) \in \mathbb{R}^2$ are independently distributed, $N_2^C = \{(z_1, z_2) \in \mathbb{R}^2 : z_1^2 + (C + z_2)^2 = C^2\}$ is a circle of radius C with the center at $(0, -C)$, and ρ_2 is Euclidean distance in $\mathbb{R}^2$.

Theorem 1 (a) establishes that the manifold S_C lies inside the set $\mathcal{M}$ bounded by an envelope we construct from a collection of p-dimensional spheres N_u. Statement (b) asserts that the distance from a point ξ to the manifold S is bounded by the distance from ξ to the furthest sphere in this collection, while (c) picks out exactly which sphere $N_{\widetilde{u}(\xi)}$ is the furthest away for a given ξ. Finally, (d) shows that the distribution of the distance from $\xi \sim N(0, I_k)$ to $N_{\widetilde{u}(\xi)}$ is the same as the distribution of the distance from a random variable η to a particular circle in $\mathbb{R}^2$ as depicted in the right panel of Figure 1.

2.3 Stochastic Bound

Theorem 1 implies a bound on the distribution of the distance from $\xi \sim N(0, I_k)$ to a p-dimensional manifold S. Assume that for some $C > 0$, S satisfies all the assumptions of Theorem 1 including Assumption 1. Then almost surely,

$$\rho^2(\xi, S) \leq \rho^2(\xi, N_{\widetilde{u}}), \tag{6}$$

as follows from statements (b) and (c) of Theorem 1. By Theorem 1 (d), the distribution of the right hand side of (6) is the same as the distribution of the random variable ψ_C

$$\psi_C = \rho_2^2(\eta, N_2^C), \tag{7}$$

where the coordinates of the two-dimensional random vector $\eta = (\sqrt{\chi_p^2}, \sqrt{\chi_{k-p}^2}) \in \mathbb{R}^2$ are independently distributed, $N_2^C = \{(z_1, z_2) \in \mathbb{R}^2 : z_1^2 + (C + z_2)^2 = C^2\}$ is a circle of radius C with the center at $(0, -C)$, and ρ_2 is Euclidean distance in $\mathbb{R}^2$. Combining these results, we establish the bound

$$P\left\{\rho^2(\xi, S) \geq x\right\} \leq P\left\{\psi_C \geq x\right\} \qquad \text{for all } x > 0,$$

so the distribution of ψ_C is an upper bound on the distribution of $\rho^2(\xi, S)$.

Note that the distribution of ψ_C depends only on the dimension of the space k, the dimension p of the manifold, and the maximal curvature $\frac{1}{C}$. The distribution of ψ_C is stochastically increasing in the maximal curvature and hence stochastically decreasing in C, so if $C_1 < C_2$ then ψ_{C_1} first-order stochastically dominates ψ_{C_2}. As $C \to \infty$ $\psi_C \Rightarrow \chi_{k-p}^2$, so if the curvature converges to zero at all relevant points then our bounding distribution converges to the distribution of the distance from $\xi \sim N(0, I_k)$ to a p-dimensional linear subspace. At the other extreme, $\psi_C \Rightarrow \chi_k^2$ as $C \to 0$ so if the curvature of the manifold becomes arbitrarily large our bound coincides with the naive bound (2) that can be imposed without any assumptions on the manifold.

We want to emphasize that what we suggest is a stochastic bound that holds under quite general assumptions. If the model of interest has additional structure, this can potentially be exploited to obtain tighter bounds.

2.4 Construction of a Feasible Test

If the manifold S satisfies the assumptions of Theorem 1 then the MD statistic is stochastically dominated by ψ_C under the null. Thus if we use $F_{1-\alpha}(C, k, p)$, the $(1-\alpha)$-quantile of ψ_C (which is easy to simulate), as a critical value the resulting test has size at most α.

A practical question is what value of C to use. According to Theorem 1, C is tied to the maximal curvature of S over the intersection of S with a cylinder D_C centered at zero. In practice, however, we do not observe the manifold S, which depends on the unknown θ_0. Nonetheless, we can see that the desired curvature is the same as the maximal curvature of the observed manifold $S^* = \{\Sigma^{-1/2}\theta(\beta), \beta \in U \subset \mathbb{R}^p\} \subset \mathbb{R}^k$ over all points in the intersection of S^* with the cylinder $D_C^*(x_0) = \{x \in \mathbb{R}^k : x - x_0 \in D_C\}$ centered at $x_0 = \Sigma^{-1/2}\theta_0$. This maximal curvature, in turn, is clearly bounded above by the maximal curvature over the whole manifold, so if we take $C^* = 1/\left(\max_{q^* \in S^*} \kappa_{q^*}(S^*)\right)$, using critical values based on ψ_{C^*} provides a test that controls size. Moreover, since C^* does not depend on any unobservables, a test based on these critical values is feasible.

3 Asymptotic Properties

The procedure described above controls finite-sample size when the reduced-form parameter estimates are normally distributed with known covariance. In practice, however, reduced-form parameter estimates $\widehat{\theta}$ are only approximately normally distributed, and researchers must estimate the covariance matrix Σ. This section obtains uniform asymptotic results, discusses the performance of our approach under conventional asymptotics, and compares our approach to others available in the literature.

3.1 Uniformity

We define a model to be a set consisting of a true value of the k-dimensional reduced-form parameter θ_0, a data generating process F_n consistent with θ_0, and a link function connecting the structural and reduced-form parameters, or more generally a manifold $\widetilde{S}_n$

describing the null hypothesis $H_0 : \theta_0 \in \widetilde{S}_n$. We assume that the null holds. We allow the data generating process F_n and the structural model $\widetilde{S}_n$ to change with the sample size n; this accommodates sequences of link functions such as those which arise under drifting asymptotic embeddings, for example the weak identification embeddings of D. Andrews and Cheng (2012) and Stock and Wright (2000). It also allows us to model the case when the researcher tries to fit a more complicated or nonlinear model when she has a larger sample. Suppose we have an estimator, $\widehat{\theta}_n$, which will be asymptotically normal with asymptotic covariance matrix $\Sigma = \Sigma(F_n)$. Let $\widehat{\Sigma}_n$ be an estimator for Σ. We consider the set of possible models $\mathcal{M} = \{M : M = (\theta_0, \{F_n\}_{n=1}^{\infty}, \{\widetilde{S}_n\}_{n=1}^{\infty})\}$ and impose the following assumption.

Assumption 2

(i) $\sqrt{n}\Sigma^{-1/2}(\widehat{\theta}_n - \theta_0) \Rightarrow N(0, I_k)$ uniformly over $\mathcal{M}$;

(ii) $\widehat{\Sigma}_n - \Sigma \to^p 0$ uniformly over $\mathcal{M}$;

(iii) The maximal and minimal eigenvalues of Σ are bounded above and away from zero uniformly over $\mathcal{M}$;

(iv) For each n and manifold $S_n = \{x = \sqrt{n}\Sigma^{-1/2}(y - \theta_0), y \in \widetilde{S}_n\}$, the manifold S_n satisfies Assumption 1 for $C = C_n = 1/\sup_{q \in S_n} \kappa_q(S_n)$.

Assumption 2(i) and (ii) require that the reduced-form parameter estimates are uniformly asymptotically normal with a uniformly consistently estimable covariance matrix. This assumption holds quite generally for many standard reduced-from estimators, such as OLS estimates and sample covariances, over large classes of models. Care is needed when using parameter estimates from ARMA models, however, as these models can suffer from near-root cancellation, leading to non-standard large-sample behavior (see D. Andrews and Cheng (2012)). Assumption 2(iii) uniformly bounds the eigenvalues of the asymptotic covariance matrix above and below, and will generally follow from a uniform bound on the moments of the data generating process. Finally, Assumption 2(iv) imposes Assumption 1. For implicitly defined manifolds, this will again follow from Lemma 1.

Description of the procedure. Let us introduce a manifold $\widehat{S}_n = \{\sqrt{n}\widehat{\Sigma}_n^{-1/2}(x - \theta_0) : x \in \widetilde{S}_n\}$, which differs from S_n in using an estimator $\widehat{\Sigma}_n$ in place of Σ. Let[4] $\widehat{C}_n =$

[4]Note that $\widehat{S}_n$ depends on θ_0 but that $\widehat{C}_n$ can also be written as $\widehat{C}_n = 1/(\sup_{q \in \widehat{S}_n^*} \kappa_q(\widehat{S}_n^*))$ for $\widehat{S}_n^* = \{\sqrt{n}\widehat{\Sigma}_n^{-1/2}(x) : x \in \widetilde{S}_n\}$, and so can be calculated.

$1/(\sup_{q\in\widehat{S}_n}\kappa_q(\widehat{S}_n))$. Our main test uses the statistic $n\min_{\theta\in\widetilde{S}_n}(\widehat{\theta}_n-\theta)'\widehat{\Sigma}_n^{-1}(\widehat{\theta}_n-\theta)$ along with critical value $F_{1-\alpha}(\widehat{C}_n,k,p)$, where we denote by $F_{1-\alpha}(C,k,p)$ the $(1-\alpha)$-quantile of the random variable ψ_C discussed in Section 2.3.

Theorem 2 *If Assumption 2 holds, then the testing procedure described above has uniform asymptotic size α:*

$$\limsup_{n\to\infty}\sup_{M\in\mathcal{M}}P\left\{n\min_{\theta\in\widetilde{S}_n}(\widehat{\theta}_n-\theta)'\widehat{\Sigma}_n^{-1}(\widehat{\theta}_n-\theta)>F_{1-\alpha}(\widehat{C}_n,k,p)\right\}\leq\alpha.$$

This result establishes the uniform asymptotic validity of our test allowing for arbitrarily nonlinear (or linear) behavior in the sequence of null hypothesis manifolds $\widetilde{S}_n$. The key to this result is that our critical values reflect the curvature of the null hypothesis manifold measured relative to the uncertainty about the reduced-form parameters for each sample size.

3.2 Curvature Under Conventional Asymptotics

Our method complements conventional asymptotic results, in the sense that if classical assumptions guaranteeing asymptotic linearity hold, then our robust critical values converge to the conventional ones. Consider a sample of size n from a model parameterized by structural parameter β that belongs to some bounded set $U\subset\mathbb{R}^p$ and assume that $\sqrt{n}(\widehat{\theta}-\theta_0)\Rightarrow N(0,\Sigma)$. Assume that the relation between structural and reduced-form parameters $\theta=\theta(\beta)$ is fixed, twice continuously differentiable with respect to β, and that the matrix $\frac{\partial}{\partial\beta}\theta(\beta)$ is full rank on a neighborhood of β_0, which is the only point in the closure of U that solves $\theta_0=\theta(\beta)$. We also assume that β_0 belongs to the interior of U. The null hypothesis manifold S_n for sample size n is the graph of function $\mathbf{x}_n(\beta)=\sqrt{n}\Sigma^{-1/2}(\theta(\beta)-\theta_0),\beta\in U$. The maximal curvature over all points of the manifold S_n is equal to $1/\sqrt{n}$ times the maximal curvature of the manifold S_1 obtained for the sample of size 1, assuming the maximal curvature is finite. Consequently, the critical value $F_{1-\alpha}(C_n,k,p)$ converges to the $(1-\alpha)$-quantile of a χ^2_{k-p}-distribution, which is the true asymptotic distribution.

3.3 Other Methods for Testing with Nuisance Parameters

The main area of application of our method is testing in the presence of nuisance parameters. There exist several alternatives. One approach widely used in practice is the

projection method, which was popularized in econometrics by Dufour and Jasiak (2001) and Dufour and Taamouti (2005). Applied in our setting, the projection method uses χ^2_k critical values and requires no assumptions beyond the asymptotic normality of $\widehat{\theta}$. However, projection method critical values are larger than those used by our method, resulting in less powerful inference in cases where our assumptions hold. Only in the limiting case of infinitely high curvature ($C = 0$) do our critical values equal those of the projection method.

An alternative to projection is to impose assumptions like those stated in Section 3.2 (typically labeled as strong identification assumptions) and use χ^2_{k-p} critical values. The obvious advantage of this approach is that it is strictly more powerful than the projection method. However, the assumptions of Section 3.2 are essential, and the test may over-reject if these assumptions fail. For example, in the Supplementary Appendix we show that weak identification may lead to significant curvature and invalidate χ^2_{k-p} critical values. By contrast, the test we suggest in this paper does not rely on such assumptions.

In the particular case of linear instrumental variables models with homoskedastic errors and multiple endogenous regressors, Guggenberger et. al. (2012) show that one may use the MD statistic with χ^2_{k-p} critical values even when the nuisance parameter may be poorly identified. By contrast, in linear IV our approach uses χ^2_k critical values.[5] However, the result of Guggenberger et. al. (2012) does not hold in more general settings. In particular, Lee (2014) provides examples of non-homoskedastic IV models in which χ^2_{k-p} critical values lead to over-rejection.

Another alternative, developed by D. Andrews and Cheng (2012), assumes we know which parameters are weakly identified and that there is a known parameter that controls the strength of identification. They then create robust tests by simulating the asymptotic distribution of the test statistic for different values of nuisance parameters and taking the "least favorable" among those distributions over a set of relevant nuisance parameter values. Unfortunately, this approach can become quite computationally demanding in models with more than a few nuisance parameters. Moreover, the assumption that a known parameter controls the strength of identification rules out many models of economic interest. As one might expect given the additional structure imposed by D. An-

[5]The null hypothesis manifold in linear IV with multiple endogenous regressors is irregular, having a singularity when the first stage is equal to zero. One can resolve this by applying the approach described below to project over the endogenous regression coefficients not under test, resulting in χ^2_k critical values.

drews and Cheng's approach, in contexts where both their results and those developed in this paper can be applied, their approach will generally yield more powerful tests. Other recent work on testing with nuisance parameters includes Elliott et al. (2015), McCloskey (2015), and Moreira and Moreira (2013).

4 Modifications and Implementation

The baseline procedure described in Section 2.4 uses the maximal curvature C^* over the whole manifold and with respect of all parameters. In this section we discuss two modifications which may allow us to further reduce the critical values. For brevity, we present these modifications under the assumption that $\widehat{\theta}$ is exactly normal with known covariance, but the asymptotic properties of these modifications are established in the Supplementary Appendix.

4.1 Modification 1: Curvature Over a Smaller Set

There are a variety of problems in which using C^* may be unappealing. For example, it may be that searching numerically for the maximal curvature over the whole manifold is quite time-consuming, or that the manifold has irregularities or points of high curvature which are far away from θ_0. In such cases we may wish to restrict attention to the curvature of the manifold over some smaller set, which raises two issues. First, we do not know the true value θ_0 and hence the center of the cylinder $D_C^*(x_0)$. Second, if the manifold is close to flat (so C is large) to find the maximal curvature over $D_C^*(x_0)$ we might need to check curvature over a huge set, which could again be very computationally demanding.

We suggest a modification which overcomes both of these problems and is easy to implement in practice. For a fixed value R, let $C \wedge R = \min\{C, R\}$. Denote by $F_{1-\alpha}(C, R, k, p)$ the $(1-\alpha)-$quantile of the distribution of $\psi_C(R)$ defined as

$$\psi_C(R) = \begin{cases} \rho_2^2(\eta, N_2^C) & \text{if } \|\eta\| \leq R; \\ \|\eta\|^2 & \text{if } \|\eta\| > R, \end{cases} \tag{8}$$

where η and and N_2^C are defined in statement (d) of Theorem 1. For any finite R the distribution of $\psi_C(R)$ first order stochastically dominates the distribution of ψ_C. In Lemma 2 below we show that one may calculate curvature only over that part of the

manifold lying inside a ball of radius proportional to R, but that one must compensate for this by using larger critical values, specifically quantiles of $\psi_C(R)$ rather than ψ_C. This is the price paid for calculating curvature over a smaller set of points.

Lemma 2 *Assume that we have a single observation $\widehat{\theta}$ from a population $\widehat{\theta} \sim N(\theta_0, \Sigma)$ with unknown mean θ_0. We wish to test the hypothesis $H_0 : \theta_0 \in \widetilde{S}$. Let $S^* = \{\Sigma^{-1/2}\theta, \theta \in \widetilde{S}\} \subset \mathbb{R}^k$ be a regular p-dimensional manifold, and $B^* = B_R(\widehat{x})$ a ball of radius $(1+\sqrt{2})R$ around $\widehat{x} = \Sigma^{-1/2}\widehat{\theta}$, where R is such that $P\{\chi^2_k \geq R^2\} < \alpha$. Let*

$$C^*_R = \begin{cases} \left(\min_{q^* \in S^* \bigcap B^*} 1/\kappa_{q^*}(S^*)\right) \wedge R, & \text{if } S^* \bigcap B^* \neq \emptyset; \\ 0, & \text{if } S^* \bigcap B^* = \emptyset. \end{cases}$$

Suppose that Assumption 1 holds for $C \wedge R$ where $C = 1/\sup_{q \in S \cap D_R} \kappa_q(S)$. Then the test which rejects the null if and only if $MD > F_{1-\alpha}(C^*_R, R, k, p)$ has size not exceeding α.

4.2 Modification 2: Working with a Subset of Parameters

The procedures discussed above treat the multi-dimensional vector β in such a way that only the direction of highest curvature affects the value of C and thus influences the critical values. Imagine instead that β can be divided into two subsets of parameters $\beta = (\beta_1', \beta_2')'$ in such a way that the curvature corresponding to directions β_1 is low. Then by calculating curvature only with respect to β_1, while projecting over β_2, we may be able to obtain smaller critical values. Moreover we can search over different partitions of β and use the one which gives us the smallest critical value. To state this result formally, let J be a subset of indexes $\{1, ..., p\}$, let β_J denote the corresponding elements of β, and let β_{-J} denote the remaining elements. Let U_{-J} and $U_J(\beta_{-J})$ denote $\{\beta_{-J} : \exists \beta_J \in \mathbb{R}^{|J|} \text{ s.t. } (\beta_J, \beta_{-J}) \in U\}$ and $\{\beta_J \in \mathbb{R}^{|J|} : (\beta_J, \beta_{-J}) \in U)\}$, respectively. Let $\mathcal{J}$ be a collection of subsets J.

Lemma 3 *Assume that $\widehat{\theta} \sim N(\theta_0, \Sigma)$, and that $S^* = \left\{\Sigma^{-1/2}\theta(\beta), \beta \in \mathbb{R}^p\right\} \subseteq \mathbb{R}^k$ is a manifold passing through θ_0. For $J \in \mathcal{J}$ and $\beta_{-J} \in U_{-J}$ consider the $|J|$-dimensional sub-manifold*

$$S^*(\beta_{-J}) = \{\Sigma^{-1/2}\theta(\beta_J, \beta_{-J}), \beta_J \in U_J(\beta_{-J})\}.$$

For $q \in S^*(\beta_{-J})$ let $\kappa_q(S^*(\beta_{-J}))$ be the curvature of the $|J|$-dimensional sub-manifold $S^*(\beta_{-J})$. Define

$$C^*_J = \inf_{\beta_{-J} \in U_{-J}} \quad \inf_{q \in S^*(\beta_{-J})} \quad \frac{1}{\kappa_q(S^*(\beta_{-J}))}$$

to be the inverse of the maximal curvature with respect to sub-parameter β_J only, where the maximum is taken over all $|J|$-dimensional sub-manifolds $S(\beta_{-J})$. Assume that for $\beta_{-J,0}$ the true value of β_{-J}, $S(\beta_{-J,0}) = \left\{x - \Sigma^{-\frac{1}{2}}\theta_0 : x \in S^*(\beta_{-J,0})\right\}$ satisfies Assumption 1 with $C = C_J^*$. Then the test which rejects the null if and only if $MD > F_{1-\alpha}(C_J^*, k, |J|)$ has size at most α. In fact, we can minimize the critical values over $\mathcal{J}$, and the test which rejects if and only if $MD > \min_{J\in\mathcal{J}} F_{1-\alpha}(C_J^*, k, |J|)$ has size at most α.

Critical values $F_{1-\alpha}(C_J^*, k, |J|)$ may be smaller than those based on the full parameter vector due to smaller curvature, or larger since $|J| \leq p$. Note, however, that so long as $\mathcal{J}$ includes the full set of indices $\{1, ..., p\}$, minimizing critical values over $\mathcal{J}$ can only decrease our critical values relative to the baseline procedure. Moreover, this modification may be freely combined with that in the previous section, allowing us to simultaneously restrict attention to a finite ball around $\widehat{\theta}$ and calculate curvature over only a subset of parameters. See Lemma S2 in the Supplement for a formal statement.

4.3 Implementation

This section summarizes how to use the results above to calculate curvature and critical values. Our discussion here will assume the manifold has an explicit global parameterization. These results may be generalized to implicitly defined manifolds, as the implicit function theorem guarantees the existence of local parameterizations at all points.[6] The null hypothesis is $H_0 : \theta_0 = \theta(\beta), \beta \in U$.

In using the first modification (Section 4.1), the appropriate choice of R will depend on the problem under study. Critical values are strictly decreasing in R, but are increasing in the maximal curvature. If the manifold S^* has singularities or points of very high curvature then it will be beneficial to choose a smaller value of R, since this gives us a better chance of excluding these points and obtaining small critical values. Likewise, in cases where the curvature optimization problem given below is computationally taxing, choosing a smaller R will reduce the domain over which we have to search. The choice of R is closely related to the choice of an initial confidence set in a Bonferroni procedure.

For the second modification (Section 4.2), from a theoretical perspective it is optimal to choose $\mathcal{J} = 2^{\{1,...,p\}}$, since search over the collection of all subsets gives the smallest

[6]Note that many implicitly defined manifolds also have an explicit representation. Thus, the assumption of a global parameterization does not preclude the application of Lemma 1

possible critical value. In high-dimensional cases, however, such a search will be unappealing. Happily, in many problems the structure of the link function θ suggests some subset of parameters which are likely to contribute to high curvature, allowing us to restrict attention to this subset. More generally, if researchers have ex-ante knowledge about which subsets of parameters tend to be problematic for conventional approaches to inference this can inform the choice of $\mathcal{J}$.

Given choices of R and $\mathcal{J}$, for each $J \in \mathcal{J}$ and $i, j \in J$ define

$$Z_J(\beta) = \Sigma^{-\frac{1}{2}} \frac{\partial}{\partial \beta_J} \theta(\beta),\ V_{J,ij}(\beta) = \Sigma^{-\frac{1}{2}} \frac{\partial^2}{\partial \beta_i \partial \beta_j} \theta(\beta)$$

$$V_{J,ij}^{\perp}(\beta) = (I - Z_J(\beta)(Z_J(\beta)' Z_J(\beta))^{-1} Z_J(\beta)') V_{J,ij}(\beta) = N_{Z_J(\beta)} V_{J,ij}(\beta).$$

We can then calculate the maximal curvature over subset J and ball $B_R(\widehat{x}) = \{x : \|x - \Sigma^{-\frac{1}{2}}\widehat{\theta}\| \leq (1+\sqrt{2})R\}$ and define

$$C_{J,R}^* = \left(\inf_{\beta \in U : \Sigma^{-\frac{1}{2}}\theta(\beta) \in B_R(\widehat{x})} \inf_{(w_1,\ldots,w_{|J|}) \in \mathbb{R}^{|J|}} \frac{\|Z_J(\beta) w\|^2}{\left\| \sum_{i,j=1}^{|J|} w_i w_j V_{ij}^{\perp}(\beta) \right\|} \right) \wedge R.$$

This is a $p + |J|$-dimensional optimization problem, which may be solved by standard techniques. We want to emphasize that no simulation is required to calculate $C_{J,R}^*$. For each J we simulate $F_{1-\alpha}(C_{J,R}^*, R, k, |J|)$ as the $1-\alpha$ quantile of random variable $\psi_C(R)$ as defined in (8) with $C = C_{J,R}^*$. If $|\mathcal{J}| > 1$ then we use the smallest critical value, $\min_{J \in \mathcal{J}} F_{1-\alpha}(C_{J,R}^*, R, k, |J|)$.

5 References

Andrews, D.W.K., and X. Cheng (2012): "Estimation and Inference with Weak, Semi-strong and Strong Identification," *Econometrica*, 80, 2153-2211.

Dufour, J.M., Jasiak, J., (2001): "Finite Sample Limited Information Inference Methods for Structural Equations and Models with Generated Regressors," *International Economic Review* 42, 815-843.

Dufour, J.M., and M. Taamouti (2005): "Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments," *Econometrica*, 73, 1351-1365.

Efron, B. (1975): "Defining the Curvature of a Statistical Problem (with Applications to Second Order Efficiency)," *Annals of Statistics*, 3, 1189-1242.

Elliott, G., U. Mueller, and M. Watson (2015): "Nearly Optimal Tests when a Nuisance Parameter is Present Under the Null Hypothesis," *Econometrica*, 83, 771 - 811.

Guggenberber P, F. Kleibergen, S. Mavroeidis, L. Chen (2012), "On the Asymptotic Sizes of subset Anderson-Rubin and Lagrance Multiplier Tests in Linear Instrumental Variables Regression,"

Econometrica, 80, 2649-2666.

Hansen L.P. and T.J. Sargent (1991): "Exact Linear Rational Expectations Models" in *Rational Expectations Econometrics*, Westview Press.

Lee, J. (2014): "Asymptotic Sizes of Subset Anderson-Rubin Tests with Weakly Identified Nuisance Parameters and General Covariance Structure," PhD Dissertation chapter, MIT.

McCloskey, A. (2015): "Bonferroni-Based Size-Correction for Nonstandard Testing Problems," unpublished manuscript.

Moreira, H. and M. Moreira (2013): "Contributions to the Theory of Optimal Tests," unpublished manuscript.

Stock, J.H., and J.H. Wright (2000): "GMM With Weak Identification," *Econometrica*, 68, 1055-1096.