Supplementary Appendix to the Paper
A Geometric Approach to Nonlinear Econometric Models

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

Abstract

This Supplementary Appendix contains supplementary material and proofs for the paper "A Geometric Approach to Nonlinear Econometric Models," by Isaiah Andrews and Anna Mikusheva. Section S1 introduces geometric concepts used in the proofs. Sections S2 and S3 prove Theorems 1 and 2 of the paper, respectively. Section S4 proves Lemma 2 from the paper and gives a related uniform asymptotic result. Section S5 proves Lemma 3 and shows that tests which both minimize critical values over subsets of parameters and restrict attention to curvature on a finite ball continue to control size. Section S6 proves Lemma 1 from the paper. Section S7 shows that models which are weakly identified in the sense of Stock and Wright (2000) imply nonlinear null hypothesis manifolds. Section S8 shows how non-linearity arises from weak identification in an analytic DSGE example. Numerical examples applying our approach to DSGE and New Keynesian Phillips Curve models may be found in the working paper version, available on Anna Mikusheva's website.[3]

First draft: May 29, 2012 This draft: November, 2015.

S1 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, while the set $V \bigcap S$ is called a coordinate neighborhood.

Note that the manifold S is defined as a set, rather than as a map. In keeping with this spirit, many of the statements below will be invariant to parametrization. We begin by developing some geometrical concepts for the special case of a regular 1-dimensional manifold, also known as a

[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.

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

curve. In particular, let S be a curve given by $\gamma : (t_0, t_1) \to \mathbb{R}^k$ where γ is twice continuously differentiable and (t_0, t_1) is an interval in $\mathbb{R}$. Let $\dot{\gamma}(t)$ and $\ddot{\gamma}(t)$ denote the first and second derivatives of γ with respect to t. Let $(\ddot{\gamma}(t))^\perp$ be the part of $\ddot{\gamma}(t)$ orthogonal to $\dot{\gamma}(t)$, then the *curvature* at $q = \gamma(t)$ is defined as $\kappa_q(S) = \frac{\|(\ddot{\gamma}(t))^\perp\|}{\|\dot{\gamma}(t)\|^2}$. One can show that this definition of curvature is invariant to parametrization. The curvature measures how quickly the curve S deviates from its tangent line local to q, and the scaling is such that a circle of radius C has curvature $1/C$ at all points.

These concepts can all be extended to general regular manifolds. Fixing a p-dimensional manifold S, for any curve $\gamma : (-\varepsilon, \varepsilon) \to S$ on S which passes through the point $q = \gamma(0) \in S$, the vector $\dot{\gamma}(0)$ is called a tangent vector to S at q. 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 $T_q(S)$). While we have defined the *tangent space* using the local coordinates $\mathbf{x}$, as one would expect $T_q(S)$ is independent of the parametrization.

To calculate the curvature at q, consider a curve $\gamma : (-\varepsilon, \varepsilon) \to S$ which lies in S and passes through $q = \gamma(0)$. Taking $T_q^\perp$ to be the $k - p$-dimensional linear space orthogonal to $T_q(S)$, define

$$\kappa_q(\gamma, S) = \frac{\|(\ddot{\gamma}(0))^\perp\|}{\|\dot{\gamma}(0)\|^2},$$

where $(W)^\perp$ stands for the projection of W onto the space $T_q^\perp$. One can show that $\kappa_q(\gamma, S)$ depends on the curve γ only through $\dot{\gamma}(0)$, so for two curves γ and γ^* in S with $\gamma(0) = \gamma^*(0) = q$ and $\dot{\gamma}(0) = \dot{\gamma}^*(0)$ we have $\kappa_q(\gamma, S) = \kappa_q(\gamma^*, S)$. We can also show that for any $X \in T_q(S)$ one can find a curve γ in S with property that $\gamma(0) = q$ and $\dot{\gamma}(0) = X$. The measure of curvature we consider is

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

This measure of curvature is closely related to the Second Fundamental Tensor (we refer the interested reader to Kobayashi and Nomizu (1969, v.2, ch. 7)), and is equal to the maximal curvature over all geodesics passing through the point q. As with the curvature measure discussed for curves, $\kappa_q(S)$ is invariant to the parametrization. Also analogous to the 1-dimensional case, if S is a p-dimensional sphere of radius C then for each $q \in S$ we have $\kappa_q(S) = 1/C$. Finally, if S is a linear subspace its curvature is zero at all points.

How to calculate curvature in practice. Let S be a p-dimensional manifold in $\mathbb{R}^k$, and let $\mathbf{x}$ be a local parametrization at a point q, $q = \mathbf{x}(y^*)$. Denote the derivatives of $\mathbf{x}$ at q by $v_i = \frac{\partial \mathbf{x}}{\partial y_i}(y^*)$. By the definition of a local parametrization, we know that the Jacobian $Z = (v_1, ..., v_p)$ is full rank, so the tangent space $T_q(S) = span\{v_1, ..., v_p\}$ is p-dimensional.

As before, for any vector $W \in \mathbb{R}^k$ let $W^\perp$ denote the part of W orthogonal to $T_q(S)$, that is, $W^\perp = N_Z W = (I - Z(Z'Z)^{-1}Z')W$. 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

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

S2 Proof of Theorem 1 of the paper

The proof is based on the following lemma:

Lemma S1 *Assume the curve $\gamma(s) : [0,b] \to D_C \subset \mathbb{R}^k$ is parameterized by arc length and that its curvature $\kappa(s) = \|\ddot{\gamma}(s)\| \le \frac{1}{C}$ for all points s. Assume that $\gamma(0) = 0$ and $\dot{\gamma}(0) = v \in$ span$\{e_1, ..., e_p\}$, where $e_1, ..., e_p$ are first p basis vectors. Then the curve $\gamma(s)$ is contained in the set $M_v \cap D_C$, where*

$$M_v = \{x : \langle x, v\rangle^2 + (C - \|x - \langle x, v\rangle v\|)^2 \ge C^2\}. \tag{S1}$$

Proof of Lemma S2. Consider the curve defined by $\beta(s) = \dot{\gamma}(s)$, the first derivative of γ. Since the curve γ is parameterized by arc length $\|\beta(s)\| = \|\dot{\gamma}(s)\| = 1$ and the new curve β lies on the unit sphere $Sph = \{x \in \mathbb{R}^k : \|x\| = 1\}$, with $\beta(0) = v$. Let $t \le \frac{\pi}{2}C$ and $t \le b$. Consider the arc length of the restriction of the curve β to the interval $[0, t]$:

$$length(t) = \int_0^t \|\dot{\beta}(s)\|ds = \int_0^t \|\ddot{\gamma}(s)\|ds = \int_0^t \kappa(s)ds \le \frac{t}{C}.$$

This implies that the geodesic (a curve of a shortest length) on the sphere Sph connecting $\beta(0)$ and $\beta(t)$ has length less than or equal to $\frac{t}{C}$ or, equivalently, that the angle between vectors $\beta(0) = v$ and $\beta(t)$ is less than or equal to $\frac{t}{C}$. Hence

$$\langle v, \beta(t)\rangle = \langle v, \dot{\gamma}(t)\rangle \ge \cos(\frac{t}{C}). \tag{S2}$$

Since $\gamma(s)$ is parameterized by arc length, from inequality (S2) we have:

$$\|\dot{\gamma}(t) - \langle v, \dot{\gamma}(t)\rangle v\| \le |\sin(\frac{t}{C})|. \tag{S3}$$

This, in turn, implies that

$$\|\gamma(t)-\langle v,\gamma(t)\rangle v\| = \|\int_0^t(\dot\gamma(s)-\langle v,\dot\gamma(s)\rangle v)ds\| \le$$
$$\le \int_0^t \|\dot\gamma(s)-\langle v,\dot\gamma(s)\rangle v\|ds \le \int_0^t \sin(\frac{s}{C})ds = C - C\cos(\frac{t}{C})$$

Inequality (S2) also implies that

$$\langle v,\gamma(t)\rangle \ge \int_0^t \cos(\frac{s}{C})ds = C\sin(\frac{t}{C}). \tag{S4}$$

Combing these results yields

$$\langle v,\gamma(t)\rangle^2 + (C - \|\gamma(t)-\langle v,\gamma(t)\rangle v\|)^2 \ge C^2$$

for all $t \le \frac{\pi}{2}C$. Notice that (S4) implies that for $\tau = \frac{\pi}{2}C$ we have $\langle v,\gamma(\tau)\rangle \ge C$ and thus for the first p coordinates of $\gamma(\tau)$, which we denote $\gamma^{(1)}(\tau)$, we have $\|\gamma^{(1)}(\tau)\| \ge C$ so the curve is leaving or has already left the cylinder D_C and thus $b \le \frac{\pi}{2}C$. This concludes the proof of the lemma. □

Proof of statement (a) of Theorem 1. First, let us show that

$$\bigcup_{\substack{v\in T_0(S)\\ \|v\|=1}} M_v = \{\|x^{(1)}\|^2 + (C - \|x^{(2)}\|)^2 \ge C^2\} = \mathcal{M}, \tag{S5}$$

where M_v is defined in (S1), $\mathcal{M}$ is defined in equation (5) of the paper and $T_0(S)$ is the tangent space to S at zero and is spanned by first p basis vectors. Indeed, the set on the left hand side consists of points x for which there exists a vector $v \in span\{e_1,...,e_p\}, \|v\| = 1$, such that

$$\langle x,v\rangle^2 + (C - \|x-\langle x,v\rangle v\|)^2 \ge C^2. \tag{S6}$$

For each x let us find the maximum of the expression on left-hand side of inequality (S6) over $v \in T_0(S), \|v\| = 1:$

$$\langle x,v\rangle^2 + (C - \|x-\langle x,v\rangle v\|)^2 =$$
$$= \langle x,v\rangle^2 + C^2 + \|x\|^2 - \langle x,v\rangle^2 - 2C\|x-\langle x,v\rangle v\| =$$
$$= C^2 + \|x\|^2 - 2C\|x-\langle x,v\rangle v\|$$

where we used that $\|x-\langle x,v\rangle v\|^2 = \|x\|^2 - \langle x,v\rangle^2$. We see that maximizing the left-hand side of

(S6) over $v \in span\{e_1, ..., e_p\}, \|v\| = 1$ is equivalent to minimizing $\|x - \langle x, v\rangle v\|$. The minimum is achieved at the projection of x onto $T_0(S) = span\{e_1, ..., e_p\}$, that is, $v = \frac{1}{\|x^{(1)}\|}(x^{(1)}, 0, ..., 0)$, where $x^{(1)} \in \mathbb{R}^p$ consists of the first p components of x. As a result, the maximum of the left-hand side of (S6) equals

$$C^2 + \|x\|^2 - 2C\|x^{(2)}\| = \|x^{(1)}\|^2 + (C - \|x^{(2)}\|)^2.$$

This proves statement (S5).

Now assume that statement (a) of Theorem 1 is incorrect and there exists a point $q \in S_C$ with $q \notin \mathcal{M}$. Take a geodesic (a curve of the shortest distance lying in S_C) $\gamma(s)$ connecting q and 0 lying in S_C, where such a curve exists since S_C is a connected manifold. Parameterize this curve by arc length. The curve $\gamma(s)$ is a geodesic in S if and only if at any point $q = \gamma(t)$ the second derivative $\ddot{\gamma}(t)$ is perpendicular to $T_q(S)$ (see Spivak (1999) for discussion of geodesics, v.3, p.3). As a result, the curvature of the geodesic γ at each point $q = \gamma(t)$ is equal to $\kappa_q(X, S)$ (where $X = \dot{\gamma}(t)$), and thus it is less than $\frac{1}{C}$. Denote the tangent to this curve at 0 by $v \in T_0(S)$. Applying Lemma S2 we obtain that the curve belongs to $M_v \cap D_C$ and thus belongs to $\mathcal{M} \bigcap D_C$. We have arrived at a contradiction. □

Proof of statement (c) of Theorem 1. Let

$$f(u) = \rho^2(\xi, N_u) = \min_{\substack{x^{(1)} \in \mathbb{R}^p, z \in \mathbb{R}_+ \\ \|x^{(1)}\|^2 + (C-z)^2 = C^2}} \|\xi^{(1)} - x^{(1)}\|^2 + \|\xi^{(2)} - zu\|^2.$$

We need to find the maximizer of $f(u)$ subject to the constraint $\|u\| = 1$. To differentiate $f(u)$ we use the "envelope theorem" that allows one to differentiate a function which is the optimum of a constrained optimization problem and yields $\frac{df(u)}{du} = -2(\xi^{(2)} - zu)$. Hence, the first-order condition for finding $\widetilde{u}$ implies that u is proportional to $\xi^{(2)}$. The sign is a reflection of the fact that we search for a max rather than a min. □

Proof of statement (b) of Theorem 1. For a given point $\xi \in \mathbb{R}^k$ find the sphere $N_{\widetilde{u}}$ furthest from ξ, where $\widetilde{u}$ is described in Theorem 1 (c), and the point $\tau \in N_{\widetilde{u}}$ such that $\rho(\xi, N_{\widetilde{u}}) = \rho(\xi, \tau)$. Consider the $k-p$ dimensional linear space $R_\tau = \{x \in \mathbb{R}^k : x^{(1)} = \tau^{(1)}\}$ that restricts the first p components of x to coincide with the first p components of τ. We prove two statements: first, that all points in the intersection $R_\tau \bigcap \mathcal{M} \bigcap D_C$ are no further from ξ than τ; and second, that this intersection $R_\tau \bigcap \mathcal{M} \bigcap D_C$ contains at least one point from S. Together, these two statements imply that $\rho(\xi, S) \leq \rho(\xi, \tau)$.

The intersection of the three sets $R_\tau \bigcap \mathcal{M} \bigcap D_C$ can be written as follows:

$$R_\tau \bigcap \mathcal{M} \bigcap D_C = \{x = (\tau^{(1)}, x^{(2)}) \in D_C : \|\tau^{(1)}\|^2 + (C - \|x^{(2)}\|)^2 \geq C^2\} =$$
$$= \left\{ x = (\tau^{(1)}, x^{(2)}) : \|x^{(2)}\| \leq C - \sqrt{C^2 - \|\tau^{(1)}\|^2} \right\}.$$

Now let us show that for each $x \in R_\tau \bigcap \mathcal{M} \bigcap D_C$ we have $\rho(\xi, x) \leq \rho(\xi, \tau)$. Indeed, one can solve the constrained maximization problem

$$\rho(\xi, x)^2 = \|\xi^{(1)} - \tau^{(1)}\|^2 + \|\xi^{(2)} - x^{(2)}\|^2 \to \max \text{ s.t. } x \in R_\tau \bigcap \mathcal{M} \bigcap D_C.$$

From the first-order condition for this problem one can see that the maximum is achieved at $x^{(2)}$ proportional to $\xi^{(2)}$. We recall that $\tau \in N_{\widetilde{u}}$ and by statement (c) $\tau^{(2)}$ is proportional to $\xi^{(2)}$. Further inspection reveals that the maximum is achieved at $x = \tau$. Hence, all points lying in the intersection $R_\tau \bigcap \mathcal{M} \bigcap D_C$ have distance to ξ less or equal than $\rho(\xi, N_{\widetilde{u}})$.

To complete the proof we need only show that $R_\tau \bigcap \mathcal{M} \bigcap D_C$ contains at least one point from the manifold S. Recall that from the definition of $\tau \in N_{\widetilde{u}}$ it follows that $\|\tau^{(1)}\| \leq C$. Then Assumption 1 guarantees that the intersection of S_C with R_τ is non-empty, while statement (a) of Theorem 1 implies that $S_C \subseteq \mathcal{M} \bigcap D_C$. □

Proof of statement (d) of Theorem 1. Note that since $\widetilde{u}$ is proportional to $\xi^{(2)}$ by statement (c), both ξ and $N_{\widetilde{u}}$ belong to the same $p+1$-dimensional linear sub-space $L_{\widetilde{u}} = \{x : x = (x^{(1)}, -z\widetilde{u}), x^{(1)} \in \mathbb{R}^p, z \in \mathbb{R}\}$. Let us restrict our attention to this subspace only. Let $(x^{(1)}, z)$ be the coordinate system in this sub-space, so ξ corresponds to $\widetilde{\xi} = (\xi^{(1)}, \|\xi^{(2)}\|)$, and $N_{\widetilde{u}}$ corresponds to the sphere $N^C = \{x = (x^{(1)}, z) \in \mathbb{R}^{p+1} : \|x^{(1)}\|^2 + (C+z)^2 = C^2\}$. The distance on $L_{\widetilde{u}}$ implied by the distance in $\mathbb{R}^k$ is the usual Euclidean metric, which we denote by $\widetilde{\rho}$. So far, we proved that $\rho(\xi, N_{\widetilde{u}}) = \widetilde{\rho}(\widetilde{\xi}, N^C)$. By invariance of the distance to orthonormal transformations of first p components we have $\widetilde{\rho}(\widetilde{\xi}, N^C) = \widetilde{\rho}(\xi^*, N^C)$, where $\xi^* = (\|\xi^{(1)}\|, 0, ..., 0, \|\xi^{(2)}\|) \in \mathbb{R}^{p+1}$. From this it is easy to see that

$$\rho(\xi, N_{\widetilde{u}}) = \rho_2(\eta, N_2^C),$$

where $\eta = (\|\xi^{(1)}\|, \|\xi^{(2)}\|) \in \mathbb{R}^2$, $N_2^C = \{(z_1, z_2) \in \mathbb{R}^2 : z_1^2 + (C+z_2)^2 = C^2\}$, and ρ_2 is Euclidian distance in $\mathbb{R}^2$. It then follows that if $\xi \sim N(0, I_k)$ then components of η have independent $\sqrt{\chi_p^2}$ and $\sqrt{\chi_{k-p}^2}$ distributions, respectively. □

S3 Proof of Theorem 2 from the paper

The procedure described in Section 2.4 of the paper guarantees that finite-sample size is controlled when the reduced-form parameter estimates are normally distributed with a known covariance matrix. In this section we prove Theorem 2 in the paper, which asserts that the procedure is asymptotically correct uniformly over a large set of models on which the reduced-form parameter estimator is uniformly asymptotically Gaussian. For ease of reference we re-state much of the discussion of Section 3.1 of the paper.

We define a model to be a set consisting of the true value of the k-dimensional reduced-form parameter θ_0, the 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. Let us 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) is the natural extension of Assumption 1 to allow for sequences of different manifolds. 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 $\widehat{C}_n = 1/(\sup_{q \in \widehat{S}_n} \kappa_q(\widehat{S}_n))$. 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:

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

This result establishes the uniform asymptotic validity of our test allowing for arbitrarily nonlinear (or linear) behavior in the sequence of null hypothesis manifolds $\widetilde{S}_n$. In particular, if curvature arises from weak identification this result allows for arbitrarily weakly or strongly identified sequences. 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.

Proof of Theorem 2. Assume that $\xi \sim N(0, I_k)$. Our main theorem states:

$$P\left\{\rho^2(\xi, S) > F_{1-\alpha}(C, k, p)\right\} \leq \alpha,$$

uniformly (over $\mathcal{M}$) for all sets $S = S_n$ if $C = C_n$ is such that the assumptions of Theorem 1 of the paper hold, that is the maximal curvature of S_n is less than $1/C_n$ and Assumption 1 is true. For the rest of the proof we suppress the index n for notational simplicity in S_n, $\widetilde{S}_n$ and $\widehat{S}_n$ and the corresponding C's. Let $\xi_n = \sqrt{n}\widehat{\Sigma}_n^{-1/2}(\widehat{\theta}_n - \theta_0)$. We note that the statistic of interest can be written as $n \min_{\theta \in \widetilde{S}}(\widehat{\theta}_n - \theta)'\widehat{\Sigma}_n^{-1}(\widehat{\theta}_n - \theta) = \rho^2(\xi_n, \widehat{S})$. Assumption 2 (i) - (iii) imply that $\xi_n \Rightarrow N(0, I_k)$ uniformly over $\mathcal{M}$.

This weak convergence can be metrized by Prokhorov's metric. Let $\beta_n \geq 0$ be Prokhorov's distance between the distributions of random variables ξ and ξ_n, where all terms are implicitly indexed by model M. According to Dudley's (1968) result we can construct a probability space and two random variables $\widetilde{\xi}_n$ and $\widetilde{\xi}$ with the same marginal distributions as ξ_n and ξ such that $P\left\{\|\widetilde{\xi} - \widetilde{\xi}_n\| > \beta_n\right\} \leq \beta_n$. From now on for simplicity of notation we will drop tildes and assume that ξ_n and ξ satisfy this condition. Thus we have:

$$P\left\{\rho(\xi_n, \widehat{S}) > F^{\frac{1}{2}}_{1-\alpha}(\widehat{C}, k, p)\right\} \leq$$
$$\leq P\left\{\rho(\xi, \widehat{S}) > F^{\frac{1}{2}}_{1-\alpha}(\widehat{C}, k, p) - \beta_n\right\} + P\left\{\|\xi - \xi_n\| > \beta_n\right\}, \tag{S7}$$

where we used that $|\rho(\xi_n, \widehat{S}) - \rho(\xi, \widehat{S})| \leq \|\xi_n - \xi\|$. The second term on the right hand side in (S7) does not exceed β_n, and β_n converges to zero uniformly over set of models $\mathcal{M}$.

Let $C = 1/(\sup_{q \in S} \kappa_q(S))$, and note that Assumption 1 from the paper holds for this value of C and the manifold S. Fix some small $\varepsilon > 0$. We can notice that:

$$P\left\{\rho(\xi, \widehat{S}) > F^{\frac{1}{2}}_{1-\alpha}(\widehat{C}, k, p) - \beta_n\right\} \leq P\{|\rho(\xi, \widehat{S}) - \rho(\xi, S)| > \varepsilon\} +$$
$$+ P\left\{\rho(\xi, S) > F^{\frac{1}{2}}_{1-\alpha}(C, k, p) - 2\varepsilon - \beta_n\right\} +$$
$$+ P\left\{|F^{\frac{1}{2}}_{1-\alpha}(\widehat{C}, k, p) - F^{\frac{1}{2}}_{1-\alpha}(C, k, p)| > \varepsilon\right\}, \tag{S8}$$

Below we show that the first and third terms on the right hand side of equation (S8) are asymptotically negligible uniformly over $\mathcal{M}$, while by choosing small ε we can bound the second term from above by a number arbitrarily close to α.

For the first term, note that the manifold $\widehat{S} = \{Ax : x \in S\}$ for matrix $A = \widehat{\Sigma}_n^{-1/2}\Sigma^{1/2}$. Let $\|X\|$ denote the matrix norm of a square matrix X (that is, the maximal eigenvalue in absolute value). Below we show that

$$|\rho(\xi, \widehat{S}) - \rho(\xi, S)| \leq 2\|\xi\| \max\{\|I - A\|, \|I - A^{-1}\|\}. \tag{S9}$$

Indeed, consider first the case when $\rho(\xi, \widehat{S}) \geq \rho(\xi, S)$, and assume that $\rho(\xi, S) = \rho(\xi, x)$ for a point $x \in S$. Then $Ax \in \widehat{S}$, and $\rho(\xi, \widehat{S}) \leq \rho(\xi, Ax)$. This implies that

$$0 \leq \rho(\xi, \widehat{S}) - \rho(\xi, S) \leq \rho(\xi, Ax) - \rho(\xi, x) \leq$$
$$\leq \rho(x, Ax) \leq \|I - A\| \cdot \|x\|$$

Next, we notice that since $0 \in S$, $\|x\| \leq \|\xi\| + \rho(\xi, x) \leq 2\|\xi\|$. The case when $\rho(\xi, \widehat{S}) < \rho(\xi, S)$ can be considered analogously. This establishes the validity of inequality (S9). Since $\|\xi\|^2$ is distributed as χ^2_k, and according to Assumptions 2 (ii)-(iii) the two maximal eigenvalues in equation (S9) converge to zero uniformly, we can see that the first term in (S8) is asymptotically small uniformly over $\mathcal{M}$.

For the second term, our main theorem guarantees that

$$P\left\{\rho(\xi,S) > F^{\frac{1}{2}}_{1-\alpha}(C,k,p) - 2\varepsilon - \beta_n\right\} \le P\left\{\psi_C^{\frac{1}{2}} > F^{\frac{1}{2}}_{1-\alpha}(C,k,p) - 2\varepsilon - \beta_n\right\}, \tag{S10}$$

where $\psi_C = \rho_2^2(\eta, N_2^C)$, for η a random vector with two independent coordinates distributed as $\left(\sqrt{\chi_p^2}, \sqrt{\chi_{k-p}^2}\right)$ and N_2^C a circle of radius C with center at point $(0,-C)$, while ρ_2 is a Euclidean distance in $\mathbb{R}^2$. First notice that function $f(x,C) = \rho_2(x, N_2^C)$ for $x \in \mathbb{R}^2$ is continuous in x uniformly over all values of C, indeed, $|\rho_2(x,N_2^C) - \rho_2(y,N_2^C)| \le \rho_2(x,y)$. Since η is continuously distributed with bounded pdf and the variable ψ_C has a pdf which is bounded above uniformly over C, this means that by choosing small enough ε and β_n we can make the right hand side in equation (S10) arbitrarily close to α.

It is easy to see that the function $\rho_2(x, N_2^C)$ is uniformly continuous in C. Thus $F_{1-\alpha}(C,k,p)$, the $(1-\alpha)$-quantile of random variable ψ_C, is continuous in C uniformly over all values of C. As $C \to \infty$ the $(1-\alpha)-$quantile $F_{1-\alpha}(C,k,p)$ converges to the $(1-\alpha)$-quantile of a χ^2_{k-p} distribution, which can be called $F_{1-\alpha}(\infty,k,p)$. For any small $\varepsilon > 0$ there exists a constant c such that for any $C > c$ we have $|F^{\frac{1}{2}}_{1-\alpha}(C,k,p) - F^{\frac{1}{2}}_{1-\alpha}(\infty,k,p)| < \varepsilon$.

What we are left to show is that (i) $\widehat{C} \to^p C$ uniformly over the subset of models $\mathcal{M}$ for which $C < c(1+\varepsilon)$ and (ii) for any probability arbitrarily close to one there exists a sample size such that for all models in $\mathcal{M}$ with $C > c(1+\varepsilon)$ we have $\widehat{C} > c$ with at least this probability.

First we examine the asymptotic relationship between C and $\widehat{C}$. Let us consider a point $q \in S$ and curvature $\kappa_q(S) = \left\|\sum_{i,j=1}^p u_i u_j V_{ij}^{\perp}\right\|$, where $(u_1,...,u_p)$ is the optimizer from formula (4) in the paper with the condition $\|\sum_{i=1}^p u_i v_i\| = 1$. By Theorem 6.4 in chapter III of Kobayashi and Nomizu (1963), there exists a unique geodesic $\gamma(t) \in S$ defined for t in an open neighborhood of zero with initial conditions: $\gamma(0) = q$ and $\dot\gamma(0) = \sum_{i=1}^p u_i v_i \in T_q(S)$. In particular, the fact that $\gamma(t)$ is a geodesic curve on S means that $\ddot\gamma(0) = \sum_{i,j=1}^p u_i u_j V_{ij}$ is perpendicular to tangent space $T_q(S)$ spanned by $Z = (v_1,...,v_p)$. This implies that $\kappa_q(S) = \|\ddot\gamma(0)\| = \|N_Z \ddot\gamma(0)\|$ and $\|\dot\gamma(0)\| = 1$.

Let us consider a curve $\widehat\gamma(t) = A\gamma(t)$ and notice that this curve lies on manifold $\widehat{S}$ and passes through the point $\widehat{q} = Aq \in \widehat{S}$. Let Z be the set of vectors spanning the tangent space $T_q(S)$, then AZ spans $T_{\widehat{q}}(\widehat{S})$.

From formula (4) of the paper we can see that

$$\kappa_{\widehat{q}}(\widehat{S}) \ge \frac{\left\|N_{AZ}\frac{d^2\widehat\gamma(0)}{dt^2}\right\|}{\left\|\frac{d\widehat\gamma(0)}{dt}\right\|^2} = \frac{\|N_{AZ}A\ddot\gamma(0)\|}{\|A\dot\gamma(0)\|^2},$$

as the left-hand-side expression is the maximum of the right-hand-side expression taken over all possible curves in $\widehat{S}$ passing through $\widehat{q}$.

Let us consider the following sequence of inequalities:

$$\kappa_q(S) = \|\ddot{\gamma}(0)\| = \|N_Z\ddot{\gamma}(0)\| \leq \|(N_{AZ}A - N_Z)\ddot{\gamma}(0)\| + \|N_{AZ}A\ddot{\gamma}(0)\| \leq$$
$$\leq \|N_{AZ}A - N_Z\|\|\ddot{\gamma}(0)\| + \kappa_{\widehat{q}}(\widehat{S})\|A\dot{\gamma}(0)\|^2. \tag{S11}$$

We can notice that $\|A\dot{\gamma}(0)\| \leq \|A\|$ since $\|\dot{\gamma}(0)\| = 1$. Finally, notice that

$$\begin{aligned} N_{AZ}A - N_Z &= A - AZ(Z'A^2Z)^{-1}Z'A^2 - I + Z(Z'Z)^{-1}Z = \\ &= (A - I) + (I - A)Z(Z'Z)^{-1}Z + AZ(Z'Z)^{-1}Z(I - A^2) + \\ &\quad + AZ(Z'Z)^{-1}Z'(A^2 - I)Z(Z'A^2Z)^{-1}Z'A^2 \end{aligned}$$

where we use A^2 to denote $A'A$. Recall that $A \to^p I$ uniformly over $\mathcal{M}$, thus $\|N_{AZ}A - N_Z\| \leq C\|I - A\|$ with probability approaching one uniformly over $\mathcal{M}$, where C is a constant that does not depend on M. Putting this reasoning together with inequality (S11) we obtain that with probability tending to one uniformly over $\mathcal{M}$

$$\kappa_q(S) \leq C\|I - A\|\kappa_q(S) + \kappa_{\widehat{q}}(\widehat{S})\|A\|^2, \tag{S12}$$

or

$$\kappa_q(S) - \kappa_{\widehat{q}}(\widehat{S}) \leq C\|I - A\|\kappa_q(S) + \kappa_{\widehat{q}}(\widehat{S})(\|A\|^2 - 1).$$

Symmetric reasoning reversing the roles of the "hatted" and "non-hatted" variables yields

$$\kappa_{\widehat{q}}(\widehat{S}) \leq C\|I - A^{-1}\|\kappa_{\widehat{q}}(\widehat{S}) + \kappa_q(S)\|A^{-1}\|^2,$$

which implies that

$$\kappa_{\widehat{q}}(\widehat{S}) - \kappa_q(S) \leq C\|I - A^{-1}\|\kappa_{\widehat{q}}(\widehat{S}) + \kappa_q(S)(\|A^{-1}\|^2 - 1),$$

and

$$\kappa_{\widehat{q}}(\widehat{S}) \leq \frac{1}{1 - C\|I - A^{-1}\|}\kappa_q(S)\|A^{-1}\|^2.$$

Since $A \to^p I$ uniformly over $\mathcal{M}$, we get that for any finite constant K, $|\kappa_{\widehat{q}}(\widehat{S}) - \kappa_q(S)| \to^p 0$ uniformly over all points $q \in S$ such that $\kappa_q(S) \leq K$ and uniformly over the set of models $\mathcal{M}$.

What we have just shown is that for any fixed constants K_1 and K_2 and $\varepsilon > 0$ we have that $|\widehat{C} - C| \to^p 0$ uniformly over all models in $\mathcal{M}$ with $K_1 < C < K_2$ and $P\{\widehat{C} > K_2(1 - \varepsilon)\} \to 1$

uniformly over all models in $\mathcal{M}$ with $C > K_2$. Inequality (S12) also implies that if $C < K_1$, that is if there exists a point $q \in S$ with $\kappa_q(S) > 1/K_1$, then $\kappa_{\widehat{q}}(\widehat{S})$ is also large and for any ε there is a sample size that guarantees $\widehat{C} < K_1(1+\varepsilon)$ with high probability for all such models uniformly over $\mathcal{M}$. Thus we have that (i) $\widehat{C} \to^p C$ uniformly over the subset of models $\mathcal{M}$ for which $C < c(1+\varepsilon)$ and (ii) for any probability arbitrarily close to one there exists a sample size such that for all models in $\mathcal{M}$ with $C > c(1+\varepsilon)$ we have $\widehat{C} > c$ with at least this probability. This concludes the proof of Theorem 2.

S4 Proof of Lemma 2 and a Related Asymptotic Result

In this section we establish two results related to the modification described in Section 4.1. First we prove Lemma 2, establishing the validity of the modified procedure which calculates curvature on a finite ball around the reduced-form parameter estimate in the exact normal model. Second, we show that this procedure has correct uniform asymptotic size under assumptions as in Section 3.1.

S4.1 Proof of Lemma 2

Proof of Lemma 2. Let $\xi = \Sigma^{-1/2}(\widehat{\theta} - \theta_0) \sim N(0, I_k)$ and $S = \{\Sigma^{-1/2}(\theta - \theta_0), \theta \in H_0\} \subset \mathbb{R}^k$. Let $\psi_C(\xi, R)$ be defined as

$$\psi_C(\xi, R) = \begin{cases} \rho^2(\xi, N_{\widetilde{u}}), & \text{if } \|\xi\| \leq R; \\ \|\xi\|^2, & \text{if } \|\xi\| > R, \end{cases}$$

where $N_{\widetilde{u}} = \{x \in \mathbb{R}^k : x = (x^{(1)}, z\widetilde{u}), x^{(1)} \in \mathbb{R}^p, z \in \mathbb{R}_+, \|x^{(1)}\|^2 + (C-z)^2 = C^2\}, \widetilde{u} = -\frac{1}{\|\xi^{(2)}\|}\xi^{(2)}$. Random variable $\psi_C(\xi, R)$ has the same distribution as $\psi_C(R)$ defined in formula (8) in the paper but is defined on a different probability space, as $\psi_C(R)$ is written in terms of the random vector $\eta \in \mathbb{R}^2$ described in Theorem 1 (d). Consider the infeasible test φ which rejects ($\varphi = 1$) if and only if $\psi_{C\wedge R}(\xi, R) \geq F_{1-\alpha}(C \wedge R, R, k, p)$. The size is $E\varphi(\xi) = \alpha$, so since $P\{\chi^2_k \geq R^2\} < \alpha$ we know that φ rejects for all realizations of ξ where $\|\xi\| > R$ as $\|\xi\| \geq \rho(\xi, N_{\widetilde{u}})$. This test is infeasible, however, since we do not know the true value of θ_0 and hence cannot calculate ξ. The (feasible) test described in Lemma 2 is

$$\widetilde{\varphi} = \begin{cases} 1, & \text{if } MD \geq F_{1-\alpha}(C^*_R, R, k, p); \\ 0, & \text{otherwise.} \end{cases}$$

We claim that $\widetilde{\varphi} \leq \varphi$ almost surely (realization-by-realization). To show that this is the case, assume that $\widetilde{\varphi} = 1$. If at the same time $\|\xi\| > R$ then $\varphi = 1$, so the claim holds. If, on the other hand, $\|\xi\| \leq R$, then the cylinder $\widetilde{D}_R(x_0)$ around $x_0 = \Sigma^{-1/2}\theta_0$ lies inside of the ball B^* of radius $(1+\sqrt{2})R$ around $\hat{x} = x_0 + \xi$, and thus

$$C_R^* = \left(\min_{q \in S^* \bigcap B^*} 1/\kappa_q(S^*)\right) \wedge R \leq \left(\min_{q \in S^* \bigcap \widetilde{D}_R(x_0)} 1/\kappa_q(S^*)\right) \wedge R \leq C. \tag{S13}$$

Indeed, to justify the last inequality, consider the two cases $R \leq C$ and $R > C$. In the first case $C_R^* \leq R \leq C$, while in the second case $\min_{q \in S^* \bigcap \widetilde{D}_R(x_0)} 1/\kappa_q(S^*) \leq C$.

Note that the function $F_{1-\alpha}(c, R, k, p)$ is decreasing in c, and hence $F_{1-\alpha}(C \wedge R, R, k, p) \leq F_{1-\alpha}(C_R^*, R, k, p)$. Further, all the assumptions of Theorem 1 are satisfied so $MD = \rho^2(\xi, S) \leq \rho^2(\xi, N_{\widetilde{u}}) \leq \psi_{C\wedge R}(\xi, R)$. Combining these results we obtain that

$$F_{1-\alpha}(C \wedge R, R, k, p) \leq F_{1-\alpha}(C_R^*, R, k, p) \leq MD = \rho^2(\xi, S) \leq \psi_{C\wedge R}(\xi, R),$$

and thus $\varphi = 1$. Hence whenever $\widetilde{\varphi} = 1$, we get that $\varphi = 1$ as well, so $\widetilde{\varphi} \leq \varphi$ as we wanted to show, and the size of the feasible test $\widetilde{\varphi}$ is bounded above by α, completing the proof. □

S4.2 Asymptotic result

Consider a set of models $\mathcal{M}$, a reduced-form parameter estimator $\widehat{\theta}_n$ and covariance estimator $\widehat{\Sigma}_n$ satisfying Assumption 2 in the paper, which is re-stated in Section S3 above. For any R such that $R^2 > \chi^2_{k,1-\alpha}$, let $\widetilde{B}_R = \{x : \|x - \widehat{\Sigma}_n^{-\frac{1}{2}}\widehat{\theta}_n\| \leq \left(1+\sqrt{2}\right)R\}$ be the ball of radius $\left(1+\sqrt{2}\right)R$ around the reduced-form parameter estimate. Let

$$\widetilde{C}_R = \begin{cases} R \wedge \left[1/\left(\max_{\widetilde{q} \in \widehat{\Sigma}_n^{-\frac{1}{2}}\tilde{S}_n \bigcap \widetilde{B}_R} \kappa_{\widetilde{q}}(\widehat{\Sigma}_n^{-\frac{1}{2}}\tilde{S})\right)\right], & \text{if } \widehat{\Sigma}_n^{-\frac{1}{2}}\tilde{S}_n \bigcap \widetilde{B}_R \neq \emptyset; \\ 0, & \text{if } \widehat{\Sigma}_n^{-\frac{1}{2}}\tilde{S}_n \bigcap \widetilde{B}_R = \emptyset. \end{cases}$$

The modified version of our 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 $F_{1-\alpha}(\widetilde{C}_R, R, k, p)$, where we denote by $F_{1-\alpha}(C_R, R, k, p)$ the $(1-\alpha)$-quantile of the random variable $\psi_C(R)$.

Theorem S1 *Under Assumption 2 with C in part (iv) replaced by $C_n \wedge R$ where $C_n = 1/\sup_{q \in S_n \cap D_R(0)} \kappa_q(S_n)$, the testing procedure described above has uniform asymptotic size α:*

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

Proof of Theorem S1. The proof of this Theorem combines the proofs of Theorem 2 and Lemma 2. For the remainder of the proof we suppress the index n for notational simplicity in S_n, $\widetilde{S}_n$ and $\widehat{S}_n$ and the corresponding C's.

We first restate the definition of C_R^*:

$$C_R^* = \begin{cases} R \wedge \left[1/\left(\max_{q^* \in S^* \bigcap B_R^*} \kappa_{q^*}(S^*)\right)\right], & \text{if } S^* \bigcap B_R^* \neq \emptyset; \\ 0, & \text{if } S^* \bigcap B_R^* = \emptyset. \end{cases}$$

where $B_R^* = \{x : \|x - \xi\| \leq (1 + \sqrt{2})R\}$. The quantity C_R^* differs from $\widetilde{C}_R$ in two respects: first, it relates to the curvature of $S^* = \Sigma^{-1/2}\widetilde{S}$, while $\widetilde{C}_R$ is connected to the curvature of $\widehat{\Sigma}_n^{-1/2}\widetilde{S}$; second the maximal curvature is found over the ball B_R^* which is centered at ξ, while $\widetilde{B}_R$ is a ball around $\widehat{\Sigma}_n^{-1/2}\widehat{\theta}_n$. Lemma 2 in the paper states that for any $r > \chi^2_{1-\alpha,k}$ and any manifold S such that the assumptions of Lemma 2 hold we have

$$P\{\rho^2(\xi, S) > F_{1-\alpha}(C_r^*, r, k, p)\} \leq \alpha. \tag{S14}$$

We proceeding along the same lines as the proof of Theorem 2 to obtain the following inequality which holds for any $r > \chi^2_{1-\alpha,k}$:

$$\begin{aligned} &P\left\{\rho(\xi_n, \widehat{S}) > F^{\frac{1}{2}}_{1-\alpha}(\widetilde{C}_R, R, k, p)\right\} \leq \\ &\leq P\left\{\rho(\xi, S) > F^{\frac{1}{2}}_{1-\alpha}(C_r^*, r, k, p) - 2\varepsilon - \beta_n\right\} + \\ &+ P\{|\rho(\xi, \widehat{S}) - \rho(\xi, S)| > \varepsilon\} + P\left\{\|\xi - \xi_n\| > \beta_n\right\} + \\ &+ P\left\{F^{\frac{1}{2}}_{1-\alpha}(\widetilde{C}_R, R, k, p) < F^{\frac{1}{2}}_{1-\alpha}(C_r^*, r, k, p) - \varepsilon\right\}. \end{aligned} \tag{S15}$$

As we argued in the proof of Theorem 2 the second and the third terms in (S15) are uniformly asymptotically negligible. Due to statement (S14) and the fact that $\psi_C(r)$ has uniformly bounded density, the first term in (S15) can be made uniformly asymptotically bounded by any value larger than α by way of choosing small $\varepsilon > 0$, since $\beta_n \to 0$ uniformly over $\mathcal{M}$. We are left only to prove that for some choice of r the last term in (S15) is uniformly asymptotically negligible. We choose $r = (1 - \delta)R$ for small $\delta > 0$.

First, we note that the distribution of random variable $\psi_C(R)$, which is defined in equation (8) in the paper, is uniformly continuous in R, thus we can always choose δ small enough that $\sup_C |F_{1-\alpha}(C, R, k, p) - F_{1-\alpha}(C, r, k, p)| < \varepsilon/2$. It is then enough to show that

$$\lim_{n\to\infty} \sup_{\mathcal{M}} P\left\{F^{\frac{1}{2}}_{1-\alpha}(\widetilde{C}_R, R, k, p) < F^{\frac{1}{2}}_{1-\alpha}(C_r^*, R, k, p) - \varepsilon/2\right\} = 0.$$

Given the monotonicity of $F_{1-\alpha}(C, R, k, p)$ it is enough to show that for any $\varepsilon_2 > 0$ (where we choose δ above so that $\delta < \varepsilon_2/R$) we have $\widetilde{C}_R \leq C_r^* + \varepsilon_2$ with probability arbitrarily close to 1 uniformly over $\mathcal{M}$ in large samples.

Let $\widetilde{q} = Aq^*$, where $A = \widehat{\Sigma}_n^{-1/2}\Sigma^{1/2}$. Note that $q^* \in S^*$ is equivalent to $\widetilde{q} \in \widehat{\Sigma}_n^{-1/2}\widetilde{S}$. Now let $\widetilde{q} \in \widetilde{B}_R$, again defined as $\|\widetilde{q} - \widehat{\Sigma}_n^{-1/2}\widehat{\theta}_n\| \leq (1+\sqrt{2})R$. We have that $\widetilde{q} - \widehat{\Sigma}_n^{-1/2}\widehat{\theta} = Aq^* - \xi_n$. Given that A uniformly converges to I and $\xi - \xi_n$ uniformly converges to zero, we have that $AB_r^* \subset \widetilde{B}_R$ with probability arbitrarily close to 1 in large samples, which in turn implies

$$\max_{\widetilde{q} \in \widehat{\Sigma}_n^{-\frac{1}{2}}\tilde{S} \bigcap \widetilde{B}_R} \kappa_{\widetilde{q}}(\widehat{\Sigma}_n^{-\frac{1}{2}}\tilde{S}) \geq \max_{\widetilde{q} \in \widehat{\Sigma}_n^{-\frac{1}{2}}\tilde{S} \bigcap AB_r^*} \kappa_{\widetilde{q}}(\widehat{\Sigma}_n^{-\frac{1}{2}}\tilde{S}). \tag{S16}$$

In the proof of Theorem 2 we showed that for $q^* \in S^*$ and $\widetilde{q} = Aq^* \in \widehat{\Sigma}_n^{-1/2}\widetilde{S}$ we have that $|\kappa_{q^*}(S^*) - \kappa_{\widetilde{q}}(\widehat{\Sigma}_n^{-1/2}\widetilde{S})|$ converges to zero uniformly over points q^* at which curvature is below a fixed constant and over $\mathcal{M}$. Hence, asymptotically (for any $\varepsilon_3 > 0$) with probability arbitrarily close to 1 we have.

$$\max_{\widetilde{q} \in \widehat{\Sigma}_n^{-\frac{1}{2}}\tilde{S} \bigcap AB_r^*} \kappa_{\widetilde{q}}(\widehat{\Sigma}_n^{-\frac{1}{2}}\tilde{S}) \geq \max_{q^* \in S^* \bigcap B_r^*} \kappa_{\widetilde{q}}(S^*) - \varepsilon_3$$

Joining this last inequality with (S16) and the definitions of C_R^* and $\widetilde{C}_R$, we arrive to the conclusion that for any positive ε_2 and any probability arbitrarily close to 1, there exists a sample size such that $\widetilde{C}_R \leq C_r^* + \varepsilon_2$ holds with at least this probability uniformly over $\mathcal{M}$. This concludes the proof of Theorem S1.

S5 Proof of Lemma 3

For ease of reference, we repeat some definitions from Section 4.2 of the paper. For 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 $\{\beta_{-J} : \exists \beta_J \in \mathbb{R}^{|J|} \ s.t. \ (\beta_J, \beta_{-J}) \in U\}$ 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

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

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.

To formalize this statement, for $J \in \mathcal{J}$ we recall the following notation from Section 4.3 of the text:

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

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

where $i, j \in J$. The inverse of the maximal curvature over subset J and ball $B_R(\widehat{x}) = \{x : \|x - \Sigma^{-\frac{1}{2}}\widehat{\theta}\| \leq (1+\sqrt{2})R\}$ is

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

Lemma 3 follows from the following Lemma, setting $R = \infty$.

Lemma S2 *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. Let $S^*(\beta_{-J})$ and $C_{J,R}^*$ be defined as above. If for all $J \in \mathcal{J}$ we have that $S^*(\beta_{-J,0})$ satisfies Assumption 1 for $C_J \wedge R$ with C_J as defined below, then the test which rejects if and only if*

$$MD > \min_{J\in\mathcal{J}} F_{1-\alpha}(C_{J,R}^*, R, k, |J|)$$

has size not exceeding α.

Proof of Lemma S2 Let $S = \{\Sigma^{-1/2}(\theta(\beta) - \theta_0), \beta \in U\} \subseteq \mathbb{R}^k$ be the infeasible manifold passing through zero. Assume that β_0 is such that $\theta(\beta_0) = \theta_0$.

Let us take any $J \in \mathcal{J}$ and consider a $|J|$-dimensional sub-manifold

$$S_J = \{\Sigma^{-1/2}(\theta(\beta_J, \beta_{-J,0}) - \theta_0), \beta_J \in U_J(\beta_{-J,0})\},$$

where $\beta_{-J,0}$ denotes the elements of the true structural parameter β_0 corresponding to indices not in set J. Let $T_0(S_J)$ be the tangent space to the manifold S_J at zero, and let $T_0^{\perp}(S_J)$ be the orthogonal complement to this space. For each $R > 0$ let us define the cylinder $D_{J,R}$ as a set of points whose orthogonal projections to $T_0(S_J)$ and $T_0^{\perp}(S_J)$ both have length at most R. Define C_J as $C_J = 1/\sup_{q \in S_J \cap D_{J,R}} \kappa_q(S_J)$.

Define $\check{J} \in \arg\min_{J \in \mathcal{J}} F_{1-\alpha}(C_J \wedge R, R, k, |J|)$, where $\check{J}$ may be selected arbitrarily when the argmin is non-unique, and let $\check{C} = C_{\check{J}} \wedge R$. Note that neither $\check{J}$ or $\check{C}$ are random. The value $F_{1-\alpha}(\check{C}, R, k, |\check{J}|)$ is an infeasible critical value which would control the size of the corresponding minimum distance test. Indeed, all the assumptions of Theorem 1 are satisfied and we have that almost surely

$$MD = \rho^2(\xi, S) \leq \rho^2(\xi, S_{\check{J}}) \leq \rho^2(\xi, N_{\check{J},\tilde{u}}) \leq \psi_{\check{J},\check{C}}(\xi, R).$$

The first inequality comes from the fact that the distance to a manifold (a set) cannot be smaller than the distance to a sub-manifold (a subset). The second inequality is the result of Theorem 1 applied to sub-manifold $S_{\check{J}}$, and the last comes defining

$$\psi_{\check{J},C}(\xi, R) = \begin{cases} \rho^2(\xi, N_{\check{J},\tilde{u}}), & \text{if } \|\xi\| \leq R; \\ \|\xi\|^2, & \text{if } \|\xi\| > R, \end{cases}$$

where $N_{J,\tilde{u}}$ is defined analogously to set N in Theorem 1 (b) and (c), re-defining $x^{(1)}$ and $x^{(2)}$ as projections on $T_0(S_{\check{J}})$ and $T_0^{\perp}(S_{\check{J}})$ respectively.

The infeasibility of the critical value $F_{1-\alpha}(\check{C}, R, k, |\check{J}|)$ comes from the fact that $\check{C}$ as well as C_J's have been calculated using infeasible (and non-random) manifold S. The remainder of the argument proceeds much as the proof of Lemma 2. In particular we notice that if realization of random variable ξ is such that $\|\xi\| > R$ then the infeasible test rejects anyway. If instead $\|\xi\| \leq R$, then the feasible critical value is almost surely (weakly) larger then the infeasible one:

$$\min_{J \in \mathcal{J}} F_{1-\alpha}(C^*_{J,R}, R, k, |J|) \geq \min_{J \in \mathcal{J}} F_{1-\alpha}(C_J \wedge R, R, k, |J|) = F_{1-\alpha}(\check{C}, R, k, |\check{J}|).$$

Indeed, repeating the proof of Lemma 2 for each $J \in \mathcal{J}$ we get an analog of formula (S13): $C^*_{J,R} \leq C_J \wedge R$, and thus due to monotonicity

$$F_{1-\alpha}(C^*_{J,R}, R, k, |J|) \geq F_{1-\alpha}(C_J \wedge R, R, k, |J|),$$

which implies the required statement. □

S6 Proof of Lemma 1

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 zero belongs to S, and in particular that $g(\mathbf{0}_k) = 0$. For some $C > 0$ let S_C denote the connected component of S intersected with D_C which contains zero. 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 Assumption 1 stated in the paper holds. In particular if $T_0(S_C)$ is spanned by the first p basis vectors then for any $y^{(1)} \in \mathbb{R}^p$ with $\|y^{(1)}\| < C$ there exists a point $x \in S_C$ with $x^{(1)} = y^{(1)}$.*

Proof of Lemma 1 Note that the implicit function theorem implies that S_C is a $p-$dimensional regular manifold. Further, S is complete by the continuity of g. Without loss of generality we assume that $T_0(S)$ is spanned by the first p basis vectors. To prove Lemma 1, we proceed by induction on the dimension p of the manifold.

Initial Step: $p = 1$ In this case the manifold S is a curve. According to the Hopf-Reinow Theorem (see e.g. Section 8.2, Theorem 5 in Bishop and Crittenden, 2001), any complete manifold is geodesically complete and, in particular, any geodesic curve that belongs to S can be indefinitely extended. Let $\gamma(t) \in S$ be a geodesic parameterized by arc length with $\gamma(0) = 0$ and $t \in [0, \infty)$. Denote by $v = \dot{\gamma}(0)$ the tangent vector at zero, which is equal the first unit vector $(1, 0, .., 0)$ up to sign. The proof of Lemma S2 implies that for $t \leq \frac{\pi}{2} C$ such that $\gamma(s) \in S_C$ $\forall s \leq t$, we have $\langle v, \gamma(t) \rangle \geq C \sin\left(\frac{t}{C}\right)$. Thus, we know that for some $\tilde{t} \leq \frac{\pi}{2} C$, $\left\langle v, \gamma\left(\tilde{t}\right)\right\rangle \geq C$. Notice that $\langle v, \gamma(t) \rangle$ is a continuous function of t and $\langle v, \gamma(0) \rangle = 0$. The intermediate value theorem gives us that for any $y^{(1)} \in [0, C]$, there exists a $t^* \in \left[0, \tilde{t}\right]$ such that the first coordinate of $\gamma(t^*)$ is $\langle v, \gamma(t^*) \rangle = y^{(1)}$. Thus, there exists a point $x \in S_C$ with $x^{(1)} = y^{(1)}$, and the result of Lemma 1 holds for $p = 1$.

Induction Step: Suppose that the conclusion of Lemma 1 holds for all $p \leq p^* - 1$. Here we prove that it holds for $p = p^*$ as well when k is held fixed and $k > p^*$.

Consider some $y^{(1)} \in \mathbb{R}^{p^*}$ with $\|y^{(1)}\| < C$, and note that $y = \left(\left(y^{(1)}\right)', \mathbf{0}_{k-p^*}\right)' \in T_0(S_C)$. Let $v \in T_0(S_C)$ be some unit vector such that $v'y = 0$. Define new function $\breve{g} : \mathbb{R}^k \to \mathbb{R}^{k-p^*+1}$ as $\breve{g}(x) = \left(g(x)', v'x\right)'$, and consider an new manifold $\breve{S} = \left\{x \in \mathbb{R}^k, \breve{g}(x) = 0\right\}$. Below we check that $\breve{S}_C$, the connected part of the new manifold laying strictly inside D_C, is a regular $p^* - 1$ dimensional manifold. In particular, Lemma S3 below states that for any $x \in S_C$ vector v is not perpendicular to the tangent space $T_x(S_C)$. Since the Jacobian of g at a point x forms a basis of the space $T_x^{\perp}(S_C)$ orthogonal to the tangent space $T_x(S_C)$, this statement implies that the Jacobian of $\breve{g}(x)$ is full rank at all $x \in S_C$. Thus, by the implicit function theorem $\check{S}_C$ is a regular $p^* - 1$ dimensional manifold, satisfying the rank condition stated in Lemma 1.

From the definition of curvature, it is easy to see that the maximal curvature of $\check{S}_C$ is less than or equal to $\frac{1}{C}$. Consequently $\check{S}_C$ is a $p^* - 1$-dimensional manifold which satisfies all the conditions of Lemma 1. By the definition of $\breve{g}$, $y \in T_0\left(\check{S}_C\right)$. Thus, by the inductive assumption there exists some $x \in \check{S}_C$ such that $x^{(1)} = y^{(1)}$. Since $\check{S}_C \subset S_C$, we have found an $x \in S_C$ such that $x^{(1)} = y^{(1)}$. Thus, Lemma 1 is proved.

Lemma S3 *Under the Assumptions of Lemma 1, for any $v \in T_0\left(S_C\right)$ with $\|v\| = 1$ and for any $x \in S_C$, we have that $v \notin T_x^{\perp}(S_C)$ where $T_x^{\perp}(S_C)$ is the linear space orthogonal to the tangent space $T_x(S_C)$.*

Proof of Lemma S3 Let $\gamma : [0, \bar{t}] \to S_C$ be a geodesic parameterized by arc-length connecting the points $\mathbf{0}_k$ and x: $\gamma(0) = 0$, $\gamma(\bar{t}) = x$ with $\bar{t} \leq \frac{\pi}{2}C$. Note that we can take $\bar{t} \leq \frac{\pi}{2}C$ since (a) we know that there exists a geodesic in S_C connecting $\mathbf{0}_k$ and x and (b) from the proof of Lemma S1 we know we can travel at most arc-length $\frac{\pi}{2}C$ along any geodesic from $\mathbf{0}_k$ before exiting the interior of D_C. The idea of the proof is to choose a unit length vector in the space $T_x(S_C)$ and, by considering parallel transport of v along the curve γ, to prove that this vector cannot be perpendicular to v. As such v cannot lie in the space orthogonal to $T_x(S_C)$.

Let $V(t) : [0, \bar{t}] \to \mathbb{R}^k$ denote the (unique) parallel transport (or translation) of vector v along curve γ. V satisfies the conditions $V(0) = v$, $V(t) \in T_{\gamma(t)}(S_C)$ and

$$\nabla_{\dot{\gamma}(t)} V(t) \equiv 0,$$

where $\nabla_{\dot{\gamma}(t)}$ denotes covariant differentiation in the direction $\dot{\gamma}(t)$. The concepts of parallel transport and covariant differentiation are discussed in most textbooks on Differential Geometry, see for example Bishop and Crittenden (2001, ch.5).

Let $II(v, w)$ be the second fundamental tensor (see Kobayashi and Nomizu (1969, v. 2, ch. 7)), then we have:

$$\frac{d}{dt} V(t) = \nabla_{\dot{\gamma}(t)} V(t) + II\left(\dot{\gamma}(t), V(t)\right) = II\left(\dot{\gamma}(t), V(t)\right).$$

Here the regular derivative $\frac{d}{dt}$ is decomposed into the covariate derivative (which belongs to the tangent space) and the part orthogonal to the tangent space, which by definition is the second fundamental form. The covariate derivative is zero since V is a parallel transport.

Next, note that for any two vectors $w, u \in T_x(S_C)$ of unit length ($\|w\| = \|u\| = 1$) we have

$$\|II(w,u)\| \le \kappa_x(S_C) \le \frac{1}{C}. \tag{S17}$$

Indeed, the second fundamental form is a bilinear transformation from $T_x(S_C) \times T_x(S_C)$ to $T_x^{\perp}(S_C)$. Let $n = \frac{1}{\|II(w,u)\|} II(w,u)$ and the consider bilinear form $h : T_x(S_C) \times T_x(S_C) \to \mathbb{R}$ defined by $h(v_1, v_2) = \langle II(v_1, v_2), n \rangle$. Any bilinear form is diagonalizable, so let v^* be the unit eigenvector corresponding to the largest eigenvalue. Then:

$$\begin{aligned} \|II(w,u)\| = h(w,u) \le h(v^*, v^*) = \langle II(v^*, v^*), n \rangle \le \\ \le \|II(v^*, v^*)\| = \kappa_x(v^*, S_C) \le \kappa_x(S_C), \end{aligned}$$

where $\kappa_x(v^*, S_C)$ is defined in equation (4) of the paper.

By the definition of parallel transport $\|V(t)\| \equiv 1$, so as we vary t, $V(t)$ traces out a curve on the unit sphere $Sph = \left\{x \in \mathbb{R}^k : \|x\| = 1\right\}$. Similar to the proof of Lemma S1, we can consider the arc-length of the curve V restricted to the interval $[0, t]$,

$$length(t) = \int_0^t \|II(\dot{\gamma}(s), V(s))\| ds \le \int_0^t \frac{1}{C} ds = \frac{t}{C}$$

where the inequality follows from (S17) applied to the vectors $\dot{\gamma}(t)$ and $V(t)$ (both belong to $T_{\gamma(t)}(S_C)$) and the assumption that maximal curvature does not exceed $\frac{1}{C}$. Thus, the angle between $V(0)$ and $V(t)$ is less than or equal to $\frac{t}{C}$, and

$$\langle V(0), V(\bar{t}) \rangle \ge \cos(\frac{\bar{t}}{C}).$$

Thus, for any $\bar{t} \in \left[0, \frac{\pi}{2} C\right)$, $\langle v, V(\bar{t}) \rangle = \langle V(0), V(\bar{t}) \rangle > 0$. Since $V(\bar{t}) \in T_x(S_C)$, however, this immediately implies that $v \notin T_x^{\perp}$, as we wanted to show. $\square$

S7 Weak Identification and Nonlinearity

In this section we note that sequences of models which are weakly identified in the sense of Stock and Wright (2000) generate asymptotically nonlinear null hypothesis manifolds. It is important to emphasize that this discussion is solely for motivation and that the validity of our method does not rely on the Stock and Wright (2000) embedding.

Consider a GMM model in which the moment function is additively separable in the data. In particular, assume that we observe a sample $\{x_i\}$ of size n consisting of identically and independently distributed observations such that

$$E(h(x_i) - \theta(\lambda, \beta)) = 0 \text{ for } \lambda = \lambda_0, \ \beta = \beta_0.$$

Here $\theta_0 = Eh(x_i)$ is a k-dimensional reduced-form parameter, while λ and β are $p_\lambda \times 1$ and $p_\beta \times 1$ vectors respectively, with $p_\lambda + p_\beta \leq k$. Assume that (λ_0, β_0) is the unique point at which the moment condition is satisfied, so the model is point identified. As in Stock and Wright (2000), we allow the function $\theta(\lambda, \beta)$ to change as the sample size grows. In particular, let

$$\theta(\lambda, \beta) = \theta_n(\lambda, \beta) = \widetilde{M}(\lambda) + \frac{1}{\sqrt{n}} M^*(\lambda, \beta),$$

where $\widetilde{M}(\lambda)$ and $M^*(\lambda, \beta)$ are fixed twice-continuously-differentiable functions with full-rank Jacobians. Stock and Wright (2000) labeled λ as strongly identified and β as weakly identified, because information about β does not accumulate as the sample size grows.

Suppose we are interested in testing hypotheses about the structural parameters λ and β. Consider first the problem of testing the hypothesis $H_0 : \beta = \beta_0$ with strongly identified nuisance parameter λ. The appropriate minimum distance test statistic is

$$MD(\beta_0) = \min_\lambda n \left(\frac{1}{n} \sum_i h(x_i) - \theta_n(\lambda, \beta_0) \right)' \Sigma^{-1} \left(\frac{1}{n} \sum_i h(x_i) - \theta_n(\lambda, \beta_0) \right),$$

where Σ is the covariance matrix of random vector $h(x_i)$ (which we take to be nonsingular) or a consistent estimate thereof. Under the null $MD(\beta_0) \Rightarrow \chi^2_{k-p_\lambda}$. Interested readers may find a full proof of this result in Stock and Wright (2000): here, we instead show that this testing problem is asymptotically equivalent to a testing problem with linear S.

Define $\xi_n = \sqrt{n}\Sigma^{-1/2}(\frac{1}{n}\sum_i h(x_i) - \theta_n(\lambda_0, \beta_0))$. By the central limit theorem, $\xi_n \Rightarrow \xi \sim N(0, I_k)$. Let the manifold S_n be the image of the function

$$\begin{aligned} m_n(\lambda) =& \sqrt{n}\Sigma^{-1/2}(\theta_n(\lambda, \beta_0) - \theta_n(\lambda_0, \beta_0)) = \\ =& \sqrt{n}\Sigma^{-1/2}(\widetilde{M}(\lambda) - \widetilde{M}(\lambda_0)) + \Sigma^{-1/2}(M^*(\lambda, \beta_0) - M^*(\lambda_0, \beta_0)) = \\ =& \sqrt{n}\Sigma^{-1/2}(\widetilde{M}(\lambda) - \widetilde{M}(\lambda_0)) + O(||\lambda - \lambda_0||). \end{aligned}$$

Then $MD(\beta_0) = \rho^2(\xi_n, S_n)$. Under standard conditions for global identification, the value of $\widetilde{M}(\lambda)$ is in a small neighborhood of $\widetilde{M}(\lambda_0)$ only if λ is close to λ_0. Under such conditions one can easily show that the range of values of λ such that $m_n(\lambda) \in S_n \bigcap \mathcal{B}$ is of order $1/\sqrt{n}$ for any bounded set $\mathcal{B}$ containing zero. Consequently, Taylor approximation shows that the intersection

$S_n \bigcap \mathcal{B}$ converges to the intersection of $\mathcal{B}$ with the p_λ-dimensional linear sub-space S spanned by the columns of the Jacobian of $\widetilde{M}(\lambda)$ at point λ_0. Informally, we may say that due to the factor $\sqrt{n}$ in the equation for m_n, as the sample size increases we zoom in on an infinitesimal neighborhood of the true value λ_0 of the strongly identified nuisance parameter. Any regular manifold, however, is arbitrarily well approximated by its tangent space on an infinitesimal neighborhood of a regular point. As a result, it is easy to show that $\rho^2(\xi_n, S_n) \Rightarrow \rho^2(\xi, S) \sim \chi^2_{k-p_\lambda}$.

Tests for hypotheses with weakly identified nuisance parameters behave quite differently. In particular, the curvature of a null hypothesis with a weakly identified nuisance parameter does not in general vanish asymptotically. To illustrate this point, assume that the hypothesis of interest is $H_0 : \lambda = \lambda_0$, so that β is a weakly identified nuisance parameter (one could equally well consider cases where the parameters of interest and nuisance parameter both contain a mix of weakly and strongly identified components: this will somewhat complicate the analysis, but will in general lead to similar conclusions). Again, we consider the appropriate minimum distance statistic:

$$MD(\lambda_0) = \min_\beta n \left(\frac{1}{n} \sum_i h(x_i) - \theta_n(\lambda_0, \beta) \right)' \Sigma^{-1} \left(\frac{1}{n} \sum_i h(x_i) - \theta_n(\lambda_0, \beta) \right).$$

Define $\xi_n = \sqrt{n}\Sigma^{-1/2}(\frac{1}{n}\sum_i h(x_i) - \theta_n(\lambda_0, \beta_0))$ as before and let S_n be the image of

$$m_n(\beta) = \sqrt{n}\Sigma^{-1/2}(\theta_n(\lambda_0, \beta) - \theta_n(\lambda_0, \beta_0)) = \Sigma^{-1/2}(M^*(\lambda_0, \beta) - M^*(\lambda_0, \beta_0)).$$

By construction, S_n is a p_β-dimensional manifold in k-dimensional Euclidean space. In contrast to the strongly identified case, however, here S_n does not change with the sample size so we may denote it by S. Hence, if S_n is nonlinear for a given sample size, it remains nonlinear in the limit. As a result, we have that

$$MD(\lambda_0) = \rho^2(\xi_n, S) \Rightarrow \rho^2(\xi, S),$$

where $\xi \sim N(0, I_k)$ and S is a p_β-dimensional manifold but is not in general a linear sub-space.

S8 DSGE Example

This section studies a highly stylized DSGE example which, unlike most DSGE models used in practice, is analytically tractable. Using this model we show that insufficiently rich dynamics for unobserved processes give rise to weak identification. We consider minimum-distance inference based on matching the auto-covariances θ of the observed series. We show that the link function has the asymptotic representation

$$\theta_n(\beta) = m(\beta_1) + \frac{1}{\sqrt{n}}\tilde{m}(\beta_1, \beta_2) + O\left(\frac{1}{n}\right).$$

where β_1 and β_2 are four- and two-dimensional transformations of the structural parameter, respectively. Thus, the structural parameter β_2 has only a small effect on the reduced-form parameter θ, and is weakly identified in the sense of Stock and Wright (2000) and thus impossible to estimate consistently. It is important to note that we do not assume this asymptotic representation, but rather derive it as a consequence of the drifting-parameter asymptotics. The key consequence from our perspective is that we can asymptotically linearize the link function with respect to β_1, but the non-linearity in β_2 remains important even in large samples, rendering classical approaches to inference inapplicable.

Assume we observe data on inflation π_t and a measure of real activity x_t for periods $t = 1, ..., n$. Suppose the dynamics of the data are described by the following small-scale model based on Clarida, Gali and Gertler (1999):

$$\begin{cases} bE_t\pi_{t+1} + \kappa x_t - \pi_t + \varepsilon_t = 0, \\ -[r_t - E_t\pi_{t+1} - \rho\Delta a_t] + E_t x_{t+1} - x_t = 0, \\ \lambda r_{t-1} + (1-\lambda)\phi_\pi \pi_t + (1-\lambda)\phi_x x_t + u_t = r_t, \end{cases} \tag{S18}$$

The first equation is a Phillips curve, the second is a linearized Euler equation and the third is a monetary policy rule. We assume that the interest rate r_t is not observed. The unobserved exogenous shocks Δa_t and u_t are generated by the following law:

$$\Delta a_t = \rho\Delta a_{t-1} + \varepsilon_{a,t}; \quad u_t = \delta u_{t-1} + \varepsilon_{u,t}; \tag{S19}$$
$$(\varepsilon_t, \varepsilon_{a,t}, \varepsilon_{u,t})' \sim iid\, N(0, \Sigma); \Sigma = diag(\sigma^2, \sigma_a^2, \sigma_u^2).$$

This is a small scale DSGE model and contains elements of many more sophisticated models used in practice. To solve the model analytically we make several further assumptions, taking $\lambda = 0, \phi_x = 0, \phi_\pi = \frac{1}{b}$ and $\sigma^2 = 0$. The model then has six unknown scalar parameters: $(b, \kappa, \rho, \delta, \sigma_u^2, \sigma_a^2)$. Under these assumptions the model (S18) is solved in Andrews and Mikusheva (2014a). The solution can be written as

$$\begin{cases} x_t = B_1 u_t + B_2\rho\Delta a_t; \\ \pi_t = \frac{\kappa}{1-\delta b} B_1 u_t + \frac{\kappa}{1-\rho b} B_2 \rho \Delta a_t, \end{cases} \tag{S20}$$

where $B_1 = -\frac{b}{b+\kappa-\delta b}$ and $B_2 = \frac{b}{b+\kappa-\rho b}$.

Andrews and Mikusheva (2014a) shows that if the persistence of two shocks is equal, $\delta = \rho$, then the model is underidentified and only four-dimensional function of the initial parameters can be uncovered from data. If $\delta = \rho + \frac{\gamma}{\sqrt{n}}$ then the model is weakly identified, where the concept of weak identification is the same as in D. Andrews and Cheng (2012).

Specifications for shock dynamics in macroeconomic models are often ad-hoc. At the same time, identification of structural parameters often requires that the dynamics of the data be sufficiently rich, which cannot be guaranteed a priori. As in the model (S18), insufficiently rich dynamics may lead to identification failure for structural parameters. As we show in Andrews and Mikusheva (2014b) for sample sizes typical in macroeconomic research a difference of 0.2 or less between ρ and δ leads to unreliable performance for conventional asymptotics in the model (S18). To study the consequences of weak identification, we consider a drifting sequence of parameter values, $\delta_n = \rho + \frac{\gamma}{\sqrt{n}}$ for bounded γ. Define:

$$A_t = \frac{b}{b+\kappa-\rho b}\rho\Delta a_t \text{ and } U_t = -\frac{b}{b+\kappa-\delta b}u_t,$$

and $\alpha = \frac{\kappa}{1-\rho b}$, $\mu = \frac{\gamma b}{1-\rho b}$. Then (S20) can be re-written as

$$\begin{cases} x_t = A_t + U_t; \\ \pi_t = \alpha A_t + \frac{\alpha}{1-\frac{\mu}{\sqrt{n}}}U_t, \end{cases}$$

where A_t and U_t are independent AR(1) processes with autoregressive coefficients ρ and $\rho + \frac{\gamma}{\sqrt{n}}$, respectively. For analytic tractability we re-parameterize from the initial $\beta = (b, \kappa, \rho, \gamma, \sigma_u^2, \sigma_a^2)$ to $\tilde{\beta} = (\rho, \gamma, \alpha, \mu, \Sigma_a, \Sigma_u)$, where $\Sigma_a = Var(A_t) = \left(\frac{b}{b+\kappa-\rho b}\right)^2 \frac{\rho^2}{1-\rho^2}\sigma_a^2$, $\Sigma_u = Var(U_t) = \left(\frac{b}{b+\kappa-\delta b}\right)^2 \frac{1}{1-\delta^2}\sigma_u^2$. This re-parametrization is one-to-one if $0 < \rho, \delta, b < 1$, $\gamma \neq 0$.

We study inference based on matching the contemporaneous covariances and first-order auto-covariances of (x_t, π_t). We choose the reduced-form parameter θ in such a way that its estimator $\widehat{\theta}$ is consistent and $\sqrt{n}$-asymptotically normal. Note that for reasonable estimators for the variance of $\widehat{\theta}$ our measure of curvature and our proposed critical values are all invariant to linear transformations of the reduced-form parameters. Further, note that while for any sample size the covariance matrix for the natural autocovariance estimator is full rank almost-surely, it is degenerate in the limit. To simplify the analysis we eliminate this degeneracy by taking a linear transformation of the reduced-form parameters, and in particular define $\alpha = \alpha_0 + \frac{\psi}{\sqrt{n}}$ and

$$y_t = \sqrt{n}(\pi_t - \alpha_0 x_t) = \psi A_t + U_t\sqrt{n}\left(\frac{\alpha_0 + \frac{\psi}{\sqrt{n}}}{1 - \frac{\mu}{\sqrt{n}}} - \alpha_0\right) =$$

$$= \psi A_t + U_t\left(\frac{\psi}{1 - \frac{\mu}{\sqrt{n}}} + \frac{\alpha_0\mu}{1 - \frac{\mu}{\sqrt{n}}}\right).$$

By the invariance of the minimum-distance test (using conventional autocovariance estimators) to linear transformation of the reduced-form parameters, the finite-sample distribution of the test derived using the linearly transformed moment condition is the same as that of the original minimum-distance statistic. This transformation is made purely to simplify the derivations, and the fact that the transformation depends on the true parameter value α_0 is irrelevant. Following this transformation, we can see that the minimum distance statistic depends on the moments:

$$\theta_n = \begin{pmatrix} Var(x_t) \\ Var(y_t) \\ cov(x_t, y_t) \\ cov(x_t, x_{t-1}) \\ cov(y_t, y_{t-1}) \\ cov(x_t, y_{t-1}) \end{pmatrix} = \begin{pmatrix} \Sigma_a + \Sigma_u \\ \psi^2\Sigma_a + \left(\frac{\psi}{1-\frac{\mu}{\sqrt{n}}} + \frac{\alpha_0\mu}{1-\frac{\mu}{\sqrt{n}}}\right)^2 \Sigma_u \\ \psi\Sigma_a + \left(\frac{\psi}{1-\frac{\mu}{\sqrt{n}}} + \frac{\alpha_0\mu}{1-\frac{\mu}{\sqrt{n}}}\right)\Sigma_u \\ \rho\Sigma_a + (\rho + \frac{\gamma}{\sqrt{n}})\Sigma_u \\ \psi^2\rho\Sigma_a + \left(\frac{\psi}{1-\frac{\mu}{\sqrt{n}}} + \frac{\alpha_0\mu}{1-\frac{\mu}{\sqrt{n}}}\right)^2 (\rho + \frac{\gamma}{\sqrt{n}})\Sigma_u \\ \psi\rho\Sigma_a + \left(\frac{\psi}{1-\frac{\mu}{\sqrt{n}}} + \frac{\alpha_0\mu}{1-\frac{\mu}{\sqrt{n}}}\right)(\rho + \frac{\gamma}{\sqrt{n}})\Sigma_u \end{pmatrix}$$

One can easily check that the natural estimate of θ_n by the corresponding sample averages is consistent and $\sqrt{n}$-asymptotically normal with a consistently estimable covariance matrix under mild parameter restrictions (e.g. ruling out unit roots). Further, the asymptotic variance matrix is bounded and positive definite provided we bound Σ_a and Σ_u above and below. Under these conditions one can show that for any sequence of constants $c_n \to \infty$,

$$\inf_{\|\tilde{\beta}\| \geq c_n} (\widehat{\theta}_n - \theta_n(\tilde{\beta}))'\widehat{\Sigma}(\widehat{\theta}_n - \theta_n(\tilde{\beta})) \to_p \infty.$$

Thus, to study the asymptotic behavior of minimum distance statistics it suffices to restrict attention to $\tilde{\beta}$ lying in bounded neighborhoods.

The link function between the reduced-form parameter θ_n and the structural parameters $\beta = (b, \kappa, \rho, \gamma, \sigma_u^2, \sigma_a^2)$ depends strongly only on a four-dimensional function of β, while the dependence on the other two directions is weak. To be exact, there exists a re-parametrization (β_1, β_2) such that β_1 and β_2 are four- and two-dimensional functions of β, respectively, and the structural model implies:

$$\theta_n(\beta) = m(\beta_1) + \frac{1}{\sqrt{n}}\tilde{m}(\beta_1, \beta_2) + O(1/n). \tag{S21}$$

Link functions of this form are often described as the Stock and Wright (2000) embedding. It is important to note, however, that we do not assume this embedding but rather find that it emerges naturally from the drifting parameter framework.

To establish (S21) note that

$$\theta_n = m(\tilde{\beta}) + \frac{1}{\sqrt{n}}\tilde{m}(\tilde{\beta}) + O(1/n),$$

uniformly over bounded neighborhoods, where

$$m(\tilde{\beta}) = \begin{pmatrix} \Sigma_a + \Sigma_u \\ \psi^2\Sigma_a + (\psi + \alpha_0\mu)^2\Sigma_u \\ \psi\Sigma_a + (\psi + \alpha_0\mu)\Sigma_u \\ \rho(\Sigma_a + \Sigma_u) \\ \rho(\psi^2\Sigma_a + (\psi + \alpha_0\mu)^2\Sigma_u) \\ \rho(\psi\Sigma_a + (\psi + \alpha_0\mu)\Sigma_u) \end{pmatrix}; \quad \tilde{m}(\tilde{\beta}) = \begin{pmatrix} 0 \\ 2(\psi + \alpha_0\mu)^2\mu\Sigma_u \\ (\psi + \alpha_0\mu)\mu\Sigma_u \\ \gamma\Sigma_u \\ (\psi + \alpha_0\mu)^2\Sigma_u(\gamma + 2\rho\mu) \\ (\psi + \alpha_0\mu)\Sigma_u(\gamma + \rho\mu) \end{pmatrix}$$

It is easy to see that $m(\tilde{\beta})$ depends only on a 4-dimensional function of $\tilde{\beta}$: $\beta_1 = (\rho, S = \Sigma_a + \Sigma_u, Z = \frac{\psi^2\Sigma_a + (\psi+\alpha_0\mu)^2\Sigma_u}{\Sigma_a+\Sigma_u}, W = \frac{\psi\Sigma_a+(\psi+\alpha_0\mu)\Sigma_u}{\Sigma_a+\Sigma_u})$. This picks out the "strongly identified" directions in the parameter space: note that we can obtain $\sqrt{n}$ consistent estimates of these parameters. However, there is a two-dimensional surface in the parameter space along which we can vary the parameters while affecting only $\tilde{m}$. We will parameterize this surface in terms of β_2, and show that the manifold obtained from $\tilde{m}(\beta_1, \beta_2)$ for fixed β_1 and different values of β_2 is non-linear.

Note that the function m does not depend on γ, so we can take this to be one of the parameters in β_2. The other parameter can be chosen as $\varsigma = \alpha_0\mu\frac{\Sigma_u}{\Sigma_a+\Sigma_u}$. In particular, the re-parametrization from $(\psi, \rho, \gamma, \mu, \Sigma_a, \Sigma_u)$ to $(\rho, S, Z, W, \gamma, \varsigma)$ is one-to-one, but the m function depends only on the first four parameters of the latter parametrization. From the definition of W we have $\psi = W - \varsigma$. From the definition of Z

$$\psi^2 + 2\varsigma\psi + \alpha_0\mu\varsigma = Z, \text{ or } \mu = \frac{Z - W^2 + \varsigma^2}{\varsigma\alpha_0}$$

and finally from the definition of ς

$$\Sigma_u = S\frac{\varsigma}{\alpha_0\mu} = S\frac{\varsigma^2}{Z - W^2 + \varsigma^2}$$

and $\Sigma_a = S - \Sigma_u$.

Re-writing $\tilde{m}$ using the new parametrization, we obtain

$$\tilde{m}(\tilde{\beta}) = \begin{pmatrix} 0 \\ 2\frac{S(Z-W^2+W\varsigma)^2}{\alpha_0\varsigma} \\ \frac{S(Z-W^2+W\varsigma)}{\alpha_0} \\ \gamma S\frac{\varsigma^2}{Z-W^2+\varsigma^2} \\ S\frac{(Z-W^2+\varsigma W)^2}{Z-W^2+\varsigma^2}\left(\gamma + 2\rho\frac{Z-W^2+\varsigma^2}{\varsigma\alpha_0}\right) \\ S\frac{(Z-W^2+\varsigma W)\varsigma}{Z-W^2+\varsigma^2}\left(\gamma + 2\rho\frac{Z-W^2+\varsigma^2}{\varsigma\alpha_0}\right) \end{pmatrix}$$

For fixed ρ, S, Z, W the two dimensional manifold obtained by allowing γ, ς to vary is non-linear.

Changes in $\beta_2 = (\gamma, \varsigma)$ produce changes in θ_n of magnitude comparable to the standard deviation of $\widehat{\theta}$, which makes it impossible consistently estimate β_2, while β_1 can be estimated precisely in large samples. In the literature it is common to call the parameter β_1 "strongly identified" in this setting, while β_2 is called "weakly identified." The key fact for us is that as the sample size increases we can linearize link function with respect to β_1 with asymptotically negligible error while the same is not true for β_2. This can be viewed as a reflection of the fact that the minimum-distance estimator of β_1 is close to the true parameter value asymptotically, and thus we can guarantee that the remainder term in a first-order Taylor expansion is asymptotically negligible. By contrast, the minimum-distance estimator of β_2 is not consistent for β_2, making Taylor approximation inaccurate. Thus, minimum-distance statistics for testing hypotheses concerning a subset of β_1 or the hypothesis of correct specification of the DSGE model will have non-standard asymptotic distributions.

This example also highlights some features of the weak identification embeddings currently studied in the literature. In the present context we can label particular functions of the parameters (β_1 and β_2) as "strongly" and "weakly" identified, respectively, but these functions relate to the original structural parameters in rather complicated ways. The "strength of identification", or more precisely the quality of conventional asymptotic approximations, depends heavily on the specific hypothesis tested. Further, existing weak identification approximations are silent about what sample size is needed to guarantee a given accuracy for conventional asymptotic approximations, even when we know that the limiting distribution of a test is standard. Finally, we can see that even in this simple highly stylized model, deriving the weakly and strongly identified directions in the parameter space is messy, and such derivations will be difficult if not impossible in richer, more empirically relevant models.

S9 References

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

Andrews, I. and A. Mikusheva (2014a): "Maximum Likelihood Inference in Weakly Identified Models," *Quantitative Economics*, forthcoming.

Andrews, I. and A. Mikusheva (2014b): "Weak Identification in Maximum Likelihood: A Question of Information," *American Economic Review: Papers and Proceedings*, 104, 195-99.

Bishop, R. and R. Crittenden (2001) "*Geometry of Manifolds*," AMS Chelsea Publishing, Providence RI.

Clarida, R., J. Gali, and M. Gertler (1999): "The Science of Monetary Policy: A New Keynesian Perspective," *Journal of Economic Literature*, 37, 1661-1707.

Dudley, R.M. (1968) "Distance of probability measures and random variables," Annals of Mathematical statistics v 39, n5, 1563-1572.

Kobayashi, S. and K. Nomizu (1969): "*Foundations of Differential Geometry*," *Interscience Publishers.*

Spivak, M. (1999): "*A Comprehensive Introduction to Differential Geometry*," *Publish or Perish, Texas.*

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