A Geometric Approach to Weakly Identified Econometric Models

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

Abstract

Many nonlinear Econometric models show evidence of weak identification. In this paper we consider minimum distance statistics and show that in a broad class of models the problem of testing under weak identification is closely related to 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, which we show can be used to detect weak identification and to construct tests robust to weak identification. We apply our new method to new Keynesian Phillips curve and DSGE examples and show that it provides a significant improvement over existing approaches.

Key words: weak identification, statistical differential geometry

This draft: October, 2013.

1 Introduction

Empirical researchers in Economics frequently find that even in large samples the data provides little information about some model parameters. In such cases, known as weakly identified, the usual asymptotic approximations to the behavior of estimators and test statistics may be quite poor, making standard approaches to inference unreliable. Weak identification has been detected in a wide range of *non-linear* estimation contexts, including estimation of the new Keynsian Phillips curve (Dufour, Khalaf, and Kichian (2006), Kleibergen and Mavroeidis (2009), Mavroeidis (2005), Nason and Smith (2008)), monetary policy rules (Mavroeidis (2010)), Dynamic Stochastic General Equilibrium (DSGE) Models (Ruge-Murcia (2007), Canova and Sala (2009), Iskrev (2010), I. Andrews and Mikusheva (2013), Guerron-Quintana, Inoue and Kilian (2013)), and Euler equations (Yogo (2004)). The need for more reliable procedures robust to weak identification in

[1]Department of Economics, M.I.T., 50 Memorial Drive, Building E52, Cambridge, MA, 02142. Email: iandrews@mit.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 Toby 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 and Jim Stock for helpful comments.

non-linear contexts has inspired a large literature in econometrics - for a survey, see Dufour (2003) and Stock, Wright, and Yogo (2002).

A number of different testing procedures have been proposed in this literature, most of which address two situations: the case in which one is interested in testing the full parameter vector, and the case in which one is interested in testing only a subset of parameters but the parameters not under test (the nuisance parameters) are strongly identified. Examples of such tests include those of Stock and Wright (2000), Guggenberger and Smith (2005, 2008), Kleibergen (2005, 2007), I. Andrews and Mikusheva (2013), and Qu (2013). The literature to date has, however, been largely silent about the case in which part of the nuisance parameter vector may be weakly identified. A notable exception is the recent paper by D. Andrews and Cheng (2012).

Our paper directly addresses the question of inference with weakly identified nuisance parameters in the context of minimum distance estimation. We suggest a fully robust testing procedure which controls size without any assumption on the strength of identification of the parameters. Further, if the nuisance parameters are strongly identified, our procedure is asymptotically equivalent to the "concentrated out" S-test suggested by Stock and Wright (2000) for hypotheses with strongly identified nuisance parameters.

Our procedure is based on novel finite-sample bounds on the distribution of the test statistic under the null, derived under the assumption that the reduced-form parameter estimate is exactly normally distributed with a known covariance matrix. We show that tests based on these bounds control asymptotic size uniformly over a broad class of models. For our asymptotic results, we assume that the reduced-form parameter estimates are uniformly asymptotically normal with consistently estimable variance. Our results hold uniformly over a large class of link functions relating the structural and reduced-form parameters and do not rely on any particular asymptotic embedding, such as those used by Stock and Wright (2000) or D. Andrews and Cheng (2012), to model weak identification.

The bounds we derive rely on techniques from differential geometry which are new in the econometrics literature. Our starting point is the observation that hypotheses in non-linear models with strongly identified nuisance parameters are asymptotically linear in a geometrical sense. In contrast, hypotheses with weakly identified nuisance parameters need not be asymptotically linear and can exhibit substantial curvature even in large samples, leading to the breakdown of the usual asymptotic approximations. Our

bounds can be viewed as a strengthening of the usual approximations, where rather than appealing to asymptotic linearity of the null hypothesis we quantify the maximal deviation of the null from linearity and use it to construct stochastic bounds. As a result, under strong-identification asymptotics our bounds recover the usual approximations. The test we suggest uses a standard minimum-distance statistic paired with easy-to-simulate robust critical values derived using these geometric bounds. The bounds we derive are also of potential interest for a range of other applications, including testing nonlinear hypotheses and inference in highly non-linear models. 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.

To date the dominant recommendation for testing hypotheses with weakly identified nuisance parameters has been the projection method (see Dufour and Jasiak (2001), Dufour and Taamouti (2005), Dufour, Khalaf, and Kichian (2006)). The strength of the projection method is that it requires no assumptions beyond the validity of the test for the full parameter vector. It is in general conservative, however, and may be extremely so in cases where the nuisance parameter is high-dimensional and/or strongly identified. Our approach is an improvement over the projection method, in that it pairs the minimum distance statistic with smaller critical values while still maintaining size.

If one knows that part of the nuisance parameter vector is strongly identified, it has been proved that in many cases one can obtain a more powerful test by concentrating out the nuisance parameter as in e.g. Stock and Wright (2000). Maintaining correct size in such cases, however, relies critically on the strong identification assumption on the nuisance parameter. In contrast, our approach requires no assumption of strong identification but, in the event that the nuisance parameters are strongly identified, is asymptotically equivalent to concentrating them out. In this sense, our robust critical values can be viewed as providing a continuous transition between projecting over and concentrating out the nuisance parameters, depending on the strength of identification.

To illustrate our approach, we revisit the question of weak identification-robust inference on new Keyensian Phillips curve parameters previously studied by Magnusson and Mavroeidis (2010). Applying our method to their empirical example we find that there is a substantial amount of curvature. Using our robust approach we construct confidence sets which both have better coverage than those developed by Magnusson and

Mavroeidis and are smaller in the empirical application. We also apply our approach to a small-scale DSGE model and find evidence of substantial curvature. We consider the problem of testing composite hypotheses about model parameters, and show that our robust critical values are substantially smaller than those used by the projection method while still controlling size.

The paper is structured as follows. In Section 2 we show that hypotheses with strongly identified nuisance parameters are asymptotically linear, while weakly identified nuisance parameters may cause non-trivial curvature of the null hypothesis. In Section 3 we derive our geometric and stochastic bounds. In Section 4 we introduce our finite-sample test as well as several modifications, under the assumption that the reduced-form parameters are normally distributed with known variance. In Section 5 we relax the finite-sample normality assumption and discuss the uniform asymptotic validity of our procedure. We also compare our testing procedures with existing methods. Section 6 presents empirical results from applying our approach to a new Keyensian Phillips curve example, and Section 7 presents simulation results from a small-scale DSGE model. Most may be found in the Appendix, while proofs of secondary importance and the details of our empirical and simulation examples are given in the Supplementary Appendix, which can be found on Anna Mikusheva's website[3].

Throughout the paper we use the following notation: $\dot{\alpha}$ is the derivative of the function α, $\ddot{\alpha}$ 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. Let $D_C = \{x = (x^{(1)}, x^{(2)}) : \|x^{(1)}\| \leq C, \|x^{(2)}\| \leq C, x^{(1)} \in \mathbb{R}^p, x^{(2)} \in \mathbb{R}^{k-p}\} \subset \mathbb{R}^k$, which is a natural generalization of a cylinder. We take N_Z to equal $I - Z(Z'Z)^{-1}Z'$.

2 Model setting

In many empirical settings researchers are interested in statistical inference on a subset of structural parameters, for example constructing confidence sets for individual parameters in multi-parameter models. Suppose we are interested in testing a hypothesis $H_0 : \alpha = \alpha_0$ about a structural parameter α but that the model also has a p-dimensional nuisance parameter (i.e. parameter not under test) β which can vary freely under the null. This paper is concerned with inference in cases where we are not willing assume that the

[3] economics.mit.edu/files/9271

nuisance parameter β is strongly identified.

We explore this issue in a minimum-distance context, and in particular assume that we have some k-dimensional ($k > p$) reduced-form parameter θ which is known to be strongly identified. Here by strong identification we mean that there is an estimator $\hat{\theta}$ of the unknown true value θ_0 such that, for some estimator $\hat{\Sigma}$ for the variance of $\hat{\theta}$, $\hat{\Sigma}^{-\frac{1}{2}}\left(\hat{\theta} - \theta_0\right)$ is uniformly asymptotically normally distributed. In time-series models θ could consist of auto-covariances or reduced-form VAR parameters, while in linear IV it could contain reduced-form regression coefficients. For the main part of our analysis we make the stronger assumption that our estimator $\hat{\theta}$ is exactly normally distributed, $\hat{\theta} - \theta_0 \sim N(0, \Sigma)$ for a known covariance matrix Σ. This allows us to develop exact finite-sample results. We then relax our assumptions to require that $\hat{\theta}$ be uniformly asymptotically normal and its asymptotic variance be uniformly consistently estimable, and show that our finite-sample results imply uniform asymptotic results.

Let $\theta(\alpha, \beta)$ be a link function connecting the structural and reduced-form parameters. For many models the link function and its derivatives can easily be calculated for any values α and β. In such cases, we consider a test based on the minimum-distance statistic

$$MD = \min_{\beta}(\hat{\theta} - \theta(\alpha_0, \beta))'\Sigma^{-1}(\hat{\theta} - \theta(\alpha_0, \beta)). \tag{1}$$

In some contexts the link function may only be defined implicitly, for example by the restriction that the true parameter values θ_0 and α_0 must satisfy $g(\theta_0, \alpha_0) = 0$ for a $(k-p)$-dimensional function g. This corresponds to having an implicit p-dimensional nuisance parameter, and in such models we define the minimum-distance statistic as

$$MD = \min_{\theta: g(\theta, \alpha_0) = 0}(\hat{\theta} - \theta)'\Sigma^{-1}(\hat{\theta} - \theta).$$

Our empirical example, inference on New Keynesian Phillips curve parameters, falls into this class.

To proceed, it is useful for us to introduce the random vector $\xi = \Sigma^{-1/2}(\hat{\theta} - \theta_0) \sim N(0, I_k)$ and the p-dimensional manifold $S = \{x : x = \Sigma^{-1/2}(\theta(\alpha_0, \beta) - \theta_0), \beta \in \mathbb{R}^p\}$. Note that under the null the manifold S is known up to a location shift determined by the true value β_0, since $\theta_0 = \theta(\alpha_0, \beta_0)$. Thus, we know the shape of S and, moreover, know that it passes through the origin. In the case of an implicitly defined link function

we have $S = \{x : x = \Sigma^{-1/2}(\theta - \theta_0), g(\theta, \alpha_0) = 0\}$. The minimum distance statistics defined above are simply the squared distance between ξ and S:

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

The distribution of $\rho^2(\xi, S)$ is in general non-standard and depends on the unknown β_0. The central statistical issue of this paper is how to define critical values such that tests based on $\rho^2(\xi, S)$ control size.

A well-known property of the normal distribution is that 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. Note that in this special case the distribution does not depend on β_0. Numerous classical results on testing in the presence of nuisance parameters are based on this fact. Indeed, 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 arbitrarily well-approximated by one in large samples. Below, we argue that testing in the presence of *strongly* identified nuisance parameters is asymptotically equivalent to testing a *linear* hypothesis, while testing in the presence of *weakly* identified nuisance parameters tends to result in asymptotically *non-linear* null hypotheses. Our main results construct bounds on the distribution of $\rho^2(\xi, S)$ and define critical values based on a measure of the non-linearity of S, specifically its maximal curvature.

There is a less informative bound that 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 $0 \in S$,

$$\rho(\xi, S)^2 = \min_{x \in S} (\xi - x)'(\xi - x) \leq (\xi - 0)'(\xi - 0) \sim \chi^2_k. \tag{3}$$

This bound is precisely the one used by the "projection method", which is currently the main approach available for testing with weakly identified nuisance parameters, see Dufour and Jasiak (2001), Dufour and Taamouti (2005), and Dufour, Khalaf, and Kichian (2006).

2.1 Strength of identification and linearity.

In this section we argue that testing with strongly identified nuisance parameters is asymptotically equivalent to the case when the manifold S is linear. In contrast, if

there are weakly identified nuisance parameters the manifold corresponding to the null hypothesis need not converge to a linear subspace, so S may remain non-linear even in large samples. To fix ideas, in this section we use the weak identification asymptotic framework introduced by Stock and Wright (2000). It is important to emphasize that the discussion in this section is solely for motivation and that the validity of our method does not rely on this or any other device for modeling weak identification.

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

$$E(h(x_i) - \theta(\alpha, \beta)) = 0 \quad \text{for} \quad \alpha = \alpha_0, \ \beta = \beta_0. \tag{4}$$

Here $\theta_0 = Eh(x_i)$ is a k-dimensional reduced-form parameter, while α and β are $p_\alpha \times 1$ and $p_\beta \times 1$ vectors respectively, with $p_\alpha + p_\beta \leq k$. Assume that (α_0, β_0) is the unique point at which the moment condition (4) is satisfied, so the model is point identified. As in Stock and Wright (2000), we allow the function $\theta(\alpha, \beta)$ to change as the sample size grows. In particular, let

$$\theta(\alpha, \beta) = \theta_n(\alpha, \beta) = \widetilde{M}(\alpha) + \frac{1}{\sqrt{n}} M^*(\alpha, \beta), \tag{5}$$

where $\widetilde{M}(\alpha)$ and $M^*(\alpha, \beta)$ are fixed twice-continuously-differentiable functions with full-rank Jacobians. In this setting α is strongly identified while β is 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 α and β. Consider first the problem of testing the hypothesis $H_0 : \beta = \beta_0$ with strongly identified nuisance parameter α. The appropriate minimum distance test statistic is

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

where Σ is the covariance matrix of random vector $h(x_i)$ (which we take to be nonsingular) or a consistent estimate thereof. Stock and Wright (2000) prove that under the null $MD(\beta_0) \Rightarrow \chi^2_{k-p_\alpha}$. 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 $\xi_n = \sqrt{n}\Sigma^{-1/2}(\frac{1}{n}\sum_i h(x_i) - \theta_n(\alpha_0, \beta_0))$. 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

$$\begin{aligned} m_n(\alpha) =& \sqrt{n}\Sigma^{-1/2}(\theta_n(\alpha, \beta_0) - \theta_n(\alpha_0, \beta_0)) = \\ =& \sqrt{n}\Sigma^{-1/2}(\widetilde{M}(\alpha) - \widetilde{M}(\alpha_0)) + \Sigma^{-1/2}(M^*(\alpha, \beta_0) - M^*(\alpha_0, \beta_0)) = \\ =& \sqrt{n}\Sigma^{-1/2}(\widetilde{M}(\alpha) - \widetilde{M}(\alpha_0)) + O(||\alpha - \alpha_0||). \end{aligned} \tag{6}$$

Then $MD(\beta_0) = \rho^2(\xi_n, S_n)$. Under standard conditions for global identification, the value of $\widetilde{M}(\alpha)$ is in a small neighborhood of $\widetilde{M}(\alpha_0)$ only if α is close to α_0. Under such conditions one can easily show that the range of values of α such that $m_n(\alpha) \in S_n \bigcap \mathcal{B}$ 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_α-dimensional linear sub-space S spanned by the columns of the Jacobian of $\widetilde{M}(\alpha)$ at point α_0. Informally, we may say that due to the factor $\sqrt{n}$ in equation (6), as the sample size increases we zoom in on an infinitesimal neighborhood of the true value α_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 $\rho^2(\xi_n, S_n) \Rightarrow \rho^2(\xi, S) \sim \chi^2_{k-p_\alpha}$, where the last step uses the fact discussed at the beginning of this section that the squared distance from a standard normal vector to a linear subspace is χ^2-distributed.

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 $H_0 : \alpha = \alpha_0$, so that β is a weakly identified nuisance parameter. Again, we consider the appropriate minimum distance statistic:

$$MD(\alpha_0) = \min_\beta n \left(\frac{1}{n}\sum_i h(x_i) - \theta_n(\alpha_0, \beta)\right)' \Sigma^{-1} \left(\frac{1}{n}\sum_i h(x_i) - \theta_n(\alpha_0, \beta)\right).$$

Define $\xi_n = \sqrt{n}\Sigma^{-1/2}(\frac{1}{n}\sum_i h(x_i) - \theta_n(\alpha_0, \beta_0))$ as before, and let S_n be the image of

$$m_n(\beta) = \sqrt{n}\Sigma^{-1/2}(\theta_n(\alpha_0, \beta) - \theta_n(\alpha_0, \beta_0)) = \Sigma^{-1/2}(M^*(\alpha_0, \beta) - M^*(\alpha_0, \beta_0)).$$

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

$$MD(\alpha_0) = \rho^2(\xi_n, S) \Rightarrow \rho^2(\xi, S),$$

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

Linearity vs strength of identification. We showed above that the problem of testing a hypothesis with strongly identified nuisance parameters is asymptotically equivalent to testing a linear hypothesis in a Gaussian model. In contrast, if there are weakly identified nuisance parameters the manifold S corresponding to the null hypothesis need not converge to a linear subspace and may remain non-linear even in large samples. Rather than focusing on strength of identification, we may view the key distinction here as between linearity and non-linearity. In particular, while strong identification guarantees that the null hypothesis will correspond to a linear subspace in the limit, even with weakly identified nuisance parameters if S (the set of parameter values satisfying the null) happens to be a linear subspace, the usual χ^2_{k-p} limiting distribution will be correct asymptotically. Hence, in models where the nuisance parameters enter the link function θ linearly, the usual (strong-identification) critical values for the minimum distance statistic will yield asymptotically valid tests regardless of the strength of identification. Asymptotic linearity, not strong identification as such, is the essential condition. On a related note, Andrews and Mikusheva (2013) show that in a parametric model, a score test which concentrates out nuisance parameters is unaffected by weak identification of these parameters provided they enter the log-likelihood function linearly.

Weak Instrumental Variable regression. A natural example of a model with nuisance parameters that enter linearly is linear IV. In particular, consider the linear IV model written in reduced-form:

$$\begin{cases} Y = \beta\pi Z + v \\ X = \pi Z + u \end{cases},$$

where Y is a $n \times 1$ vector of realizations of the dependent variable, X is an endogenous regressor of interest, and Z is a set of k instrumental variables. The hypothesis of interest is $H_0 : \beta = \beta_0$, where the nuisance parameter π takes values in $\mathbb{R}^k$.

To construct a minimum-distance statistic we match reduced-form coefficients. Let $\theta(\beta, \pi) = (\beta\pi', \pi')'$ be the link function between the $2k \times 1$-dimensional reduced form parameter θ and structural parameters β and π. We take $\hat{\theta} = vec((Z'Z)^{-1}Z'(Y, X))$ to be the OLS estimator of the reduced-form coefficients. Under mild assumptions $\hat{\theta}$ is consistent and $\sqrt{n}$-asymptotically normal with consistently estimable variance matrix Σ. Algebraic manipulation shows that the minimum distance statistic in this case is a version of the well-known Anderson-Rubin (1949) statistic. The null manifold $S = \{\sqrt{n}\Sigma^{-1/2}(\theta(\beta_0, \pi) - \theta_0), \pi \in \mathbb{R}^k\}$ is a k-dimensional linear subspace in $\mathbb{R}^{2k}$. Thus, by the argument above we can use χ^2_k critical values. Hence, in this case our approach gives us the Anderson-Rubin test, which is known to be robust to weak instruments. This example happens to be trivial from a theoretical perspective, as the nuisance parameter π enters linearly.

3 Geometry

In the previous section we argued that the distribution of the statistic MD depends on the shape of the null hypothesis manifold S. In this section, we begin by introducing a number of geometrical concepts. Using these tools, and in particular the curvature of S, we bound MD from above for each realization of the data. From this realization-by-realization bound we then derive an upper bound on the distribution of MD that depends only on the dimension and curvature of S.

3.1 Manifolds, tangent spaces, curvature

In this paper we focus on regular manifolds embedded in k-dimensional Euclidean space with the usual Euclidean norm $\|\cdot\|$. 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 $\mathbf{x} : U \to V \bigcap S$ from an open set $U \subset \mathbb{R}^p$ 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. Hence, if we have different parameterizations for the same manifold, which of them we use is entirely a matter of convenience. In some problems it may be the case that there does not exist a global parametrization, that is a fixed mapping $\mathbf{x}$ satisfying the conditions above such that S is the image of $\mathbf{x}$. For instance, suppose that as discussed in Section 2, S is defined as the set of points q satisfying the $k-p$-dimensional restriction $g(q) = 0$. If we assume that the twice continuously differentiable function g has a Jacobian of rank $k-p$ at all points, the Implicit Function Theorem guarantees the existence of a local parametrization in a neighborhood of each point $q \in S$ but a global parametrization need not exist.

We begin by developing some geometrical concepts for the special case of a regular 1-dimensional manifold, also known as a curve. In particular, let S be a curve given by $\alpha : (t_0, t_1) \to \mathbb{R}^k$ where α is twice continuously differentiable and (t_0, t_1) is an interval in $\mathbb{R}$. The *arc length* is defined as $s(t) = \int_{t_0}^{t} \|\dot{\alpha}(\tau)\| d\tau$. Without loss of generality, we can take α to be parameterized by arc length s, in which case at all points $\|\dot{\alpha}(s)\| = 1$ and the vector $\ddot{\alpha}(s)$ is perpendicular to $\dot{\alpha}(s)$. The vector $\dot{\alpha}(s)$ is called the tangent vector to S at $q = \alpha(s)$, while $\kappa_q(S) = \|\ddot{\alpha}(s)\|$ is called the *curvature* at q. 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.

The change of variables from the arbitrary parametrization t to arc length s is not necessary for the calculation of curvature. In particular, as before let $\dot{\alpha}(t)$ and $\ddot{\alpha}(t)$ denote the first and second derivatives of α, now with respect to t. If we let $(\ddot{\alpha}(t))^{\perp}$ be the part of $\ddot{\alpha}(t)$ orthogonal to $\dot{\alpha}(t)$, then the curvature at $q = \alpha(t)$ is $\kappa_q(S) = \frac{\|(\ddot{\alpha}(t))^{\perp}\|}{\|\dot{\alpha}(t)\|^2}$. One can show that this definition of curvature is invariant to parametrization, and hence that in the special case of a curve parameterized by arc length it reduces to the definition given above.

These concepts can all be extended to general regular manifolds. Fixing a p-dimensional manifold S, for any curve $\alpha : (-\varepsilon, \varepsilon) \to S$ on S which passes through the point $q = \alpha(0) \in S$ the tangent vector $\dot{\alpha}(0)$ 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 from its geometrical interpretation $T_q(S)$ is independent of the parametrization.

To calculate the curvature at q, consider a curve $\alpha : (t_0, t_1) \to S$ which lies in S and passes through $q = \alpha(0)$. Taking $T_q^\perp$ to be the $k-p$-dimensional linear space orthogonal to $T_q(S)$, define

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

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

$$\kappa_q(S) = \sup_{X \in T_q(S), \dot{\alpha}(0) = X} \kappa_q(\alpha, S) = \sup_{X \in T_q(S), \dot{\alpha}(0) = X} \frac{\left\|(\ddot{\alpha}(0))^\perp\right\|}{\|\dot{\alpha}(0)\|^2}. \tag{7}$$

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 curves passing through the point q. As with the curvature measure discussed for curves, (7) 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

$$\kappa_q(S) = \sup_{\substack{u=(u_1,..,u_p)\in\mathbb{R}^p \\ \|\sum_{i=1}^p u_i v_i\|=1}} \left\| \sum_{i,j=1}^{p} u_i u_j V_{ij}^{\perp} \right\| = \sup_{(w_1,...,w_p)\in\mathbb{R}^p} \frac{\left\| \sum_{i,j=1}^{p} w_i w_j V_{ij}^{\perp} \right\|}{\left\| \sum_{i=1}^{p} w_i v_i \right\|^2}. \tag{8}$$

Notice that here we calculate maximal curvature over all directions in the tangent space; the possibility of calculating curvature over only a subset of directions is discussed in Section 4.2.

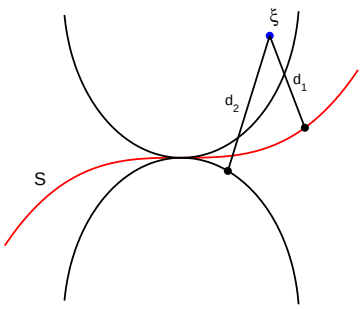
3.2 Geometric bounds

In this section we establish a bound on 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 $\kappa_q(S)$ over all relevant points in the manifold S. Our bound will be based on global properties of the manifold, in the sense of properties that hold on a fixed bounded set, but we abstract from the behavior of the manifold at infinity as 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 walk the reader through 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 the curve lies between them- see Figure 1 for illustration. Since S lies between the circles, the distance from any point ξ to S (denoted by d_1 in 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

Figure 1: Bounding a line between two circles.



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 or a manifold of co-dimension 1.[4] If a regular $k-1$ dimensional manifold S in $\mathbb{R}^k$ has curvature $\kappa_q(S) \leq 1/C$ at all $q \in S$, consider the two $k-1$-dimensional spheres of radius C that are tangent to the manifold at zero. One can show (see Theorem 1 below) that as in the one-dimensional case S lies between these two spheres. Hence, we again have that the distance from any point ξ to S is bounded above by the distance from ξ to the furthest of the two spheres. Likewise, if the maximal curvature of S goes to zero (so that $C \to \infty$) we again have that on any bounded set the two spheres converge to the tangent space to S at zero, which in this case is a $k-1$-dimensional hyperplane.

Dealing with manifolds of co-dimension greater than 1 is much 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 $\mathbb{R}^3$ which 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 left panel of Figure 2: as in the case of co-dimension 1, we can 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 Figure 2) is bounded above by the

[4] The co-dimension of a manifold is the difference between the dimension of the space and the dimension of the manifold.

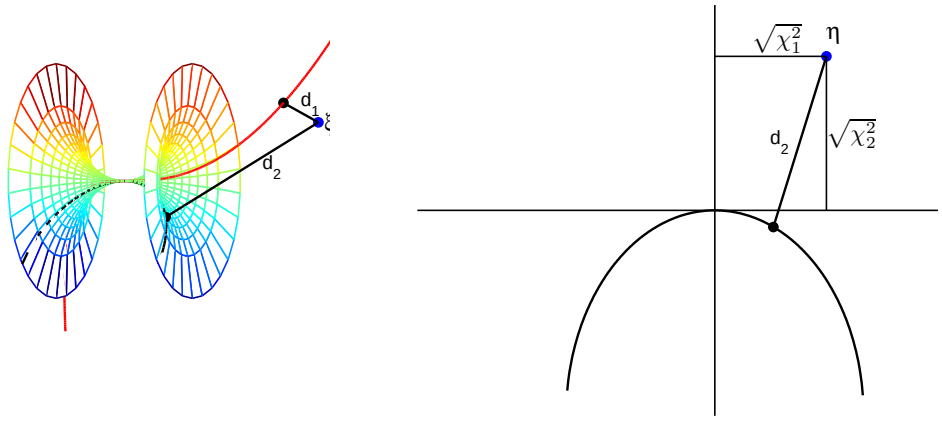


Figure 2: Left panel: the envelope for a space curve in $\mathbb{R}^3$. Right panel: distribution of the distance d_2

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 we consider 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 depicted on the right panel of Figure 2 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) if $S \bigcap D_C$ is not connected. 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 $x^{(1)} = y^{(1)}$.*

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. We have already imposed a local dimensionality assumption on S by restricting the rank of the tangent space at all points. The distribution of the minimum distance statistic, however, depends on global properties of the manifold S and so to bound the distribution we need a global dimensionality assumption. To illustrate why local dimensionality assumptions are insufficient, imagine a strip $S = \{(x, y, z) \in \mathbb{R}^3 : z = 0, -\varepsilon < y < \varepsilon, x \in \mathbb{R}\}$ in $\mathbb{R}^3$. At any point $q \in S$ the dimension of the tangent space is equal to 2, but if $\varepsilon > 0$ is small enough then S does not satisfy Assumption 1. For ε sufficiently small, however, the distance from ξ to S behaves like the distance from ξ to the line $S^* = \{(x, y, z) : y = 0, z = 0\}$, which is one dimensional both locally and globally.

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

$$\mathcal{M} = \{\|x^{(1)}\|^2 + (C - \|x^{(2)}\|)^2 \geq C^2\}. \tag{9}$$

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

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

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

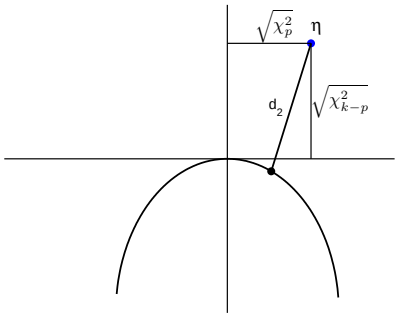
(c) $\max_{u \in \mathbb{R}^{p-k}, \|u\|=1} \rho(\xi, N_u) = \rho(\xi, N_{\widetilde{u}})$, where $\widetilde{u} = -\frac{1}{\|\xi^{(2)}\|}\xi^{(2)}$.

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

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

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 Euclidian distance in $\mathbb{R}^2$.

Figure 3: The stochastic bound described in Theorem 1 (d).



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 Figure 3.

Theorem 1 (b) relies on Assumption 1. In some models, like our empirical example in Section 6, this assumption holds trivially. In other contexts, we may try to check Assumption 1 numerically. In particular, we can draw points at random from the manifold S, for example running MCMC on the restricted model. We can then project these draws on the tangent space. As the number of draws increases, we can check that the sampler draws points in all parts of the ball of radius C in the tangent space. This bears a resemblance to how one would check that a Gibbs sampling procedure visits the whole parameter space.

3.3 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,

$$\rho^2(\xi, S) \leq \rho^2(\xi, N_{\tilde{u}}), \tag{10}$$

as follows from statements (b) and (c) of Theorem 1. By Theorem 1 (d), the distribution of the right hand side of (10) is the same as the distribution of the random variable ψ_C defined as

$$\psi_C = \rho_2^2(\eta, N_2^C), \tag{11}$$

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 $\rho^2(\xi, S)$. We make the following observations:

(1) The distribution of ψ_C depends only on the dimension of the space k, the dimension of the manifold p and the maximal value of the curvature, $\frac{1}{C}$.

(2) 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}.

(3) $\psi_C \Rightarrow \chi_{k-p}^2$ as $C \to \infty$, 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.

(4) 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 (3) 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.

4 Statistical application of the stochastic bound and its modifications

In this section, we discuss how we can use the bounds obtained above to construct tests based on MD which control size in finite samples under the assumption that $\hat{\theta} - \theta_0$ is normally distributed with known variance Σ. In this section we also propose three modifications of our baseline procedure that offer practical advantages in some contexts and which, as we discuss below, may be freely combined to suit the preferences of the researcher and the circumstances at hand.

As we note in equation (2) the minimum distance statistic is equal to $\rho^2(\xi, S)$, where $\xi \sim N(0, I_k)$ and the manifold $S = \{\Sigma^{-1/2}(\theta(\beta) - \theta_0), \beta \in U \subset \mathbb{R}^p\} \subset \mathbb{R}^k$ describes the restrictions imposed on the reduced-form parameters by the null, and passes through zero if the null is true. If the manifold S satisfies the assumptions of Theorem 1 then by the results of Section 3.3 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 not exceeding 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.

If the null hypothesis has a global parametrization, as when $H_0 : \theta_0 = \theta(\beta)$, $\beta \in U$,

let $\kappa(\beta) = \kappa_{q^*=\Sigma^{-1/2}\theta(\beta)}(S^*)$. The latter is a function on U and depends only on the first two derivatives of $\theta(\beta)$. Hence, if we can evaluate these derivatives, finding $C^* = 1/\left(\max_{\beta\in U}\kappa(\beta)\right)$ is a standard non-stochastic optimization problem. If $\theta(\beta)$ is fairly tractable we may be able to solve for C^* analytically, while if not we can use the usual menu of numerical optimization techniques, such as Newton's method.

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 $\widehat{\theta}$. 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 the curvature over a huge set, which could again be very computationally demanding.

We suggest a test 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 $F_\alpha(C, R, k, p)$ the $\alpha-$quantile of the distribution of $\psi_C(R)$ defined as

$$\psi_C(R) = \begin{cases} \rho_2^2(\eta, N_2^C) & \text{if } \|\eta\| \le R; \\ \|\eta\|^2 & \text{if } \|\eta\| > R, \end{cases} \tag{12}$$

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 1 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 $\psi_C(R)$ rather than ψ_C. This is the price paid for calculating curvature over a smaller set of points.

Lemma 1 *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 $H_0 : \theta_0 = \theta(\beta)$ for some $\beta \in U \subset \mathbb{R}^p$. Let $S^* = \{\Sigma^{-1/2}\theta(\beta), \beta \in U \subset \mathbb{R}^p\} \subset \mathbb{R}^k$ 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*

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

Assume that for any $x \in S^*$ *such that* $\|x - \widehat{x}\| \leq R$ *we have that the projection of* $S^* \bigcap B_R(x)$ *onto* $T_x(S^*)$ *contains a* p*-dimensional ball centered at* x *with radius* $C^*_R \wedge R$. *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

Suppose we wish to test a hypothesis of the form $H_0 : \theta_0 = \theta(\beta)$ for some value of the p-dimensional structural parameter $\beta \in \mathbb{R}^p$. 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 sub-sets of parameters $\beta = (\beta_1', \beta_2')'$ in such a way that the curvature corresponding to directions β_1 is high, but the null hypothesis exhibits only a low degree of curvature in the directions corresponding to β_2. Let p_1 be the dimension of β_1, p_2 the dimension of β_2, and $p = p_1 + p_2$. In this section we propose a modified procedure that treats β_1 and β_2 differently. In particular we may reduce the critical value of the test due to the low curvature with respect to β_2 while projecting over β_1. The modified procedure may be more powerful if the difference in curvature with respect to parameters β_1 and β_2 is large.

For any value β_1 consider the p_2-dimensional manifold $S(\beta_1) = \{\Sigma^{-1/2}\theta(\beta_1, \beta_2), \beta_2 \in \mathbb{R}^{p_2}\}$. For any point $q = \Sigma^{-1/2}\theta(\beta_1, \beta_2) \in S(\beta_1)$ find the curvature $\kappa_q(S(\beta_1))$. Let

$$C_{\beta_2} = \min_{\beta_1} \min_{q \in S(\beta_1)} 1/\kappa_q(S(\beta_1)), \tag{13}$$

be the inverse of the maximal curvature with respect to β_2, where the maximum is taken over all p_2-dimensional sub-manifolds $S(\beta_1)$. One also needs that the analog of Assumption 1 is satisfied, specifically that for the true value $\beta_{1,0}$ the projection of $S(\beta_{1,0})$ onto its tangent space at zero covers a p_2-dimensional ball of radius C_{β_2}. The test which rejects the null if and only if $MD > F_{1-\alpha}(C_{\beta_2}, k, p_2)$ controls size. Note that the

new critical value may increase since $p_2 < p$, but it may also decrease if the curvature corresponding to directions β_2 is substantially smaller than the maximal curvature over all directions.

If we calculate curvature multiple times, projecting over any collection of different subsets of β and using the smallest critical value, the resulting test will still control size. Thus, it is perfectly valid to search over all subsets of the parameter vector and choose the one which, when we project over it, yields the smallest critical value. Moreover, we can combine projection with the approach discussed in Section 4.1 and calculate the maximal curvature of $S(\beta_1)$ only on a ball of radius $(1+\sqrt{2})R$ around the reduced-form parameter estimate, and the resulting tests will again control size.

4.3 Modification 3: Pre-test for weak identification

If for some reason a researcher prefers not to use our robust critical values, we propose a simple "pre-test of weak identification" which is of a similar flavor to procedures based on the first-stage F-statistic in linear IV. Specifically, imagine that a researcher wants to use a robust procedure (for example our robust critical values or the projection method) unless she knows that nonlinearity will not cause large size distortions, in which case she prefers to use standard, non-robust critical values. Our bounds can be used to address this question, and in particular to determine whether nonlinearity constitutes a problem in a given setting. Below we suggest an approach which, used as the first step in a two-step test, ensures that the procedure as a whole controls size.

To proceed, let us introduce the notion of a "tolerance level". Suppose we would like to have a test of size α, but are uncertain whether the usual strong-identification asymptotics provide a reasonable approximation in our context; in the event that these approximations are imperfect, we are willing to accept a test with true size $\alpha + \alpha^*$ in exchange for the additional power and convenience of using conventional critical values. The potential size distortion α^* is called the tolerance level and was previously discussed by e.g. Stock and Yogo (2005).

The pretest we propose asks whether the curvature of the model is sufficiently small to ensure that tests based on classical χ^2_{k-p} critical values (that is, tests which treat the nuisance parameter β as well-identified) with nominal size α have true size not exceeding $\alpha + \alpha^*$. To determine whether this is the case we calculate C^*, the smallest value C

such that the $(1-\alpha)$-quantile of a χ^2_{k-p} distribution does not exceed $F_{1-(\alpha+\alpha^*)}(C,k,p)$. The cut-off C^* depends on the dimension k of the reduced-form parameter vector and the dimension p of the nuisance parameter. To implement the pre-test we can calculate the value C over the whole manifold and compare C to C^*. If $C > C^*$ the researcher can safely concentrate out β and use χ^2_{k-p} critical values while if $C \leq C^*$ she should use a robust procedure. We can guarantee that the resulting two-step test will have size less than $\alpha + \alpha^*$. Table S1 in the Supplementary Appendix reports the cut-offs C^* for nominal 5% tests with tolerance level 5% for different values of p and k. We can see that for a fixed dimension k of the reduced-form parameter, increasing the number of nuisance parameters p tightens the restrictions imposed on curvature if one wants to concentrate out the nuisance parameters. Rather than calculating the maximal curvature over the whole manifold, or over all parameters, we can also create pre-tests based on the modified robust critical values discussed in Sections 4.1 and 4.2.

The main difference between the pre-testing procedures we suggest and the majority of existing tests for weak identification is that our approach guarantees that the true size does not exceed the pre-specified tolerance level. In contrast, many alternative approaches to detecting identification failure in nonlinear models, for example those discussed by Inoue and Rossi (2011), Iskrev (2010) and Wright (2003), do not control the size distortion of two-step testing procedures.

5 Asymptotic procedure

5.1 Uniformity result

The main procedure described above 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 show that the procedure is asymptotically correct uniformly over a large set of models for which the reduced-form parameter estimator is asymptotically Gaussian.

We take a model to be a set consisting of the true value of the k-dimensional reduced-form parameter θ_0, its estimator $\widehat{\theta}$, the true value of uncertainty associated with this estimate (asymptotic covariance matrix Σ), an estimator $\widehat{\Sigma}$, and a link function connecting the structural and reduced form parameters, or more generally a manifold $\widetilde{S}$

describing the null hypothesis. The hypothesis of interest is $H_0 : \theta \in \widetilde{S}$. Here by estimators $\widehat{\theta}$ and $\widehat{\Sigma}$ we understand some procedures or algorithms for producing such estimators, where $\widehat{\theta}_n$ and $\widehat{\Sigma}_n$ are estimators for the sample of size n. We allow $\widetilde{S}$ to change with the sample size n as well, and will denote it $\widetilde{S}_n$. This allows for sequences of link functions such as those which arise under drifting asymptotic embeddings like those of Andrews and Cheng (2012) and Stock and Wright (2000). We consider a set of models $\mathcal{M} = \{M : M = (\theta_0, \widehat{\theta}, \Sigma, \widehat{\Sigma}, \widetilde{S})\}$ satisfying the following assumptions.

Assumption 2

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

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

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

(iv) For each n the manifold $S_n = \{x = \sqrt{n}\Sigma^{-1/2}(y - \theta_0), y \in \widetilde{S}_n\}$ belongs to the set $\mathcal{S}$ described below.

Assumption 3 Let $\mathcal{S}$ be a set of regular p-dimensional manifolds in $\mathbb{R}^k$ passing through zero where for each $S \in \mathcal{S}$ there is some $C(S)$ such that for S_C as defined in Section 3.2, $C(S) = 1/(\sup_{q \in S_{C(S)}} \kappa_q(S))$ and the orthogonal projection of S_C onto the tangent space $T_0(S)$ covers the ball of radius C centered at zero.

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 $F_\alpha(C, k, p)$ the α-quantile of the random variable ψ_C discussed in Section 3.3.

Theorem 2 *The testing procedure described above has uniform asymptotic size α:*

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

In the Supplementary Appendix we show that the test which maximizes curvature over the ball of radius $(1+\sqrt{2})R$ around $\hat{\Sigma}^{-\frac{1}{2}}\hat{\theta}$ (the modification suggested in Section 4.1) likewise has correct uniform asymptotic size over $\mathcal{M}$, and establishing analogous results for the other modifications discussed in the previous section is straightforward.

5.2 Curvature of strongly identified parameters.

In Section 2.1 we motivated our geometric perspective by showing that if a set of nuisance parameters is strongly identified in the sense of Stock and Wright (2000), the null hypothesis is asymptotically linear in these parameters. Here we show that the curvature corresponding to strongly identified parameters is asymptotically of order $O(1/\sqrt{n})$, where n is the sample size.

Consider a sample of size n from a model parameterized by structural parameter β that belongs to some bounded set $U \subseteq \mathbb{R}^p$ and assume that one has consistent and asymptotically normal estimates for the reduced-form parameter θ:

$$\sqrt{n}(\hat{\theta} - \theta_0) \Rightarrow N(0, \Sigma).$$

Assume that the relation between structural and reduced-form parameters $\theta = \theta(\beta)$ is fixed (not changing with n), twice continuously differentiable with respect to β, and that the matrix $\frac{\partial}{\partial\beta}\theta(\beta)$ has full rank in a neighborhood of β_0, which is the only point in the closure of U that solves the equation $\theta_0 = \theta(\beta)$. 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. This can easily be seen by examining the role of the scale of parametrization $\mathbf{x}_n$ in formula (8). As a result, the inverse curvature C_n for the sample of size n is proportional to $\sqrt{n}$ and so diverges to infinity. Consequently, the critical value $F_{1-\alpha}(C_n, k, p)$ converges to the $(1-\alpha)$-quantile of the χ^2_{k-p}-distribution, which is the true asymptotic distribution under the assumption of the strong identification.

To summarize, our procedure makes no assumption about the strength of identification of the nuisance parameter but if the nuisance parameter happens to be strongly identified our robust critical values converge to those derived under the assumption of strong identification. Thus, our baseline approach to constructing robust critical values does not sacrifice power in cases with strongly identified nuisance parameters.

5.3 Comparison with other methods available for testing hypotheses with weak nuisance parameters

As previously discussed, there is a wide literature discussing weak-identification-robust tests for the full parameter vector and for hypotheses with strongly identified nuisance parameters, but much less is known about testing with weakly identified nuisance parameters.

Projection method. The projection method has recently been the standard approach to inference with weakly identified nuisance parameters. The projection method was popularized in econometrics by Dufour and Jasiak (2001) and Dufour and Taamouti (2005), and recent applications to non-standard testing problems in econometrics include Dufour, Khalaf, and Kichian (2006), Guerron-Quintana, Inoue and Kilian (2013), and Qu (2013).

The projection method is based on the observation that the hypothesis $H_0 : \alpha = \alpha_0$ with nuisance parameter β is equivalent to the hypothesis

$$H_0 : \exists \ \ \beta_0 \text{ s.t. } \alpha = \alpha_0, \beta = \beta_0.$$

Let $MD(\alpha_0, \beta_0) = n(\widehat{\theta} - \theta(\alpha_0, \beta_0))'\widehat{\Sigma}^{-1}(\widehat{\theta} - \theta(\alpha_0, \beta_0))$ be the test statistic for the hypothesis $H_0 : \alpha = \alpha_0, \beta = \beta_0$, and note that under the assumptions of Theorem 2 it is uniformly asymptotically χ^2_k under the null. Recall that the minimum distance statistic for testing a hypothesis on α alone is $MD(\alpha_0) = \inf_{\beta^*} MD(\alpha_0, \beta^*)$, so since

$$MD(\alpha_0) = \inf_{\beta^*} MD(\alpha_0, \beta^*) \leq MD(\alpha_0, \beta_0) \Rightarrow \chi^2_k,$$

we know that comparing $MD(\alpha_0)$ to χ^2_k critical values will yield a test which controls size. The name "projection method" stems from the fact that constructing confidence sets for α based on this procedure is equivalent to constructing a joint confidence set for (α_0, β_0) using the full-vector MD statistic and then projecting this set on the parameter space for α.

The obvious advantage of the projection method is that it requires no assumption about the strength of identification of β, since it relies only on the validity of the test for the full parameter vector. Other advantages include its ease of implementation and broad

applicability. The primary disadvantage of the projection method is its conservativeness. Our test, introduced in Section 4, is based on the same statistic as the projection method but uses smaller critical values while still maintaining size. Only in the limiting case of infinitely high curvature ($C = 0$) do our critical values equal those of the projection method. As a result, except for this limiting case our test is strictly more powerful than the projection method and produces smaller confidence sets in all realizations of the sample.

Concentrating out nuisance parameters. If one knows that the nuisance parameter β in a given testing problem is strongly identified then he/she can simply "concentrate out" the nuisance parameter and pair the $MD(\alpha_0)$ statistic with χ^2_{k-p} critical values. As discussed in Section 2.1, this reduction in degrees of freedom (from k to $k-p$) stems from the fact that any manifold corresponding to a hypothesis with a strongly identified nuisance parameter converges to a linear subspace asymptotically.

The obvious advantage of this approach is that it is strictly more powerful than the projection method. However, the assumption of strong identification of the nuisance parameter is essential, and the test may over-reject if this assumption fails. In many practical settings, including the DSGE and Phillips curve examples discussed below, the exact nature and source of weak identification is not clear, and we are unaware of any previous test of the null of weak identification which can be used to separate the weakly and strongly identified parameters. In contrast, the test we suggest in this paper does not employ any assumptions about the strength of identification of any parameter in point identified models. Indeed, since our approach is based on a finite-sample perspective we do not even require that there be a meaningful distinction between weakly and strongly identified structural parameters in the model.

Other methods. There are very few other papers that work directly with weakly identified nuisance parameters. One alternative approach, developed by D. Andrews and Cheng (2012), proceeds by assuming we know the structure of weak identification. In particular, they assume we know which parameters are weakly identified and that there is a known parameter that controls the strength of identification. Under these assumptions, they show that statistics for testing hypotheses with weakly identified nuisance parameters have non-standard asymptotic distributions which depend on the value

of the nuisance parameter. They then propose creating robust tests by simulating the asymptotic distribution of the test statistic for different values of the nuisance parameter 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, for example the new Keyensian Phillips curve and DSGE examples discussed below. As one might expect given the additional structure imposed by D. Andrews and Cheng's approach, in contexts where both their results and those developed in this paper can be applied, their approach will generally yield smaller critical values.

Another example of inference with weakly identified nuisance parameters is given in I. Andrews and Mikusheva (2013). That paper considers a case when concentrating out a weakly identified parameter leads to asymptotically correct inferences, but this result holds only for weakly identified parameters which enter the log-likelihood function linearly.

6 Empirical application: new Keynesian Phillips curve

To illustrate the methods developed in this paper we consider the application of our approach to an empirical example drawn from Magnusson and Mavroeidis (2010) (henceforth MM), who study the problem of identification-robust inference on new Keynesian Phillips curve (NKPC) parameters. These parameters govern the behavior of inflation and play a central role in monetary policy-making. There is substantial evidence that NKPC parameters are weakly identified: see Mavroeidis, Plagborg-Moller and Stock (2013) for extensive discussion of this issue.

6.1 Model setting

MM consider identification-robust inference on new Keynesian Phillips curve parameters, building on the minimum-distance approach of Sbordone (2005). They study the simple NKPC model

$$\pi_t = \frac{(1-\nu)^2}{\nu(1+\varrho)} x_t + \frac{1}{1+\varrho} E\left[\pi_{t+1} | \mathcal{I}_t\right] + \frac{\varrho}{1+\varrho}\pi_{t-1} + \varepsilon_t \tag{14}$$

where x_t is a measure of marginal costs, π_t is inflation, $E[\cdot|\mathcal{I}_t]$ denotes an expectation conditional on information available at time t, and ε_t is an exogenous shock with $E[\varepsilon_{t+1}|\mathcal{I}_t] = 0$. The structural parameters ν and ϱ denote the degree of price stickiness and price indexation, respectively. MM further assume that (π_t, x_t) follows a third order VAR process which can be written in companion form as

$$z_t = A(\theta)z_{t-1} + \epsilon_t,$$

where $z_t = (\pi_t, x_t, \pi_{t-1}, x_{t-1}, \pi_{t-2}, x_{t-2})'$ is a 6×1 vector, $A(\theta)$ is a 6×6 companion form matrix, the reduced-form parameter θ is a vector of 12 unknown VAR coefficients, and ϵ_t are VAR innovations with $E[\epsilon_{t+1}|\mathcal{I}_t] = 0$. MM note that the NKPC model (14) implies that the true parameter value (θ, ν, ϱ) satisfies the 6-dimensional restriction

$$f(\theta, \nu, \varrho) = A(\theta)' \left\{ \left[I - \frac{1}{1+\varrho} A(\theta)' \right] e_\pi - \frac{(1-\nu)^2}{\nu(1+\varrho)} e_x \right\} - \frac{\varrho}{1+\varrho} e_\pi = 0,$$

where e_π and e_x are unit vectors with $e_\pi' z_t = \pi_t$ and $e_x' z_t = x_t$

Under mild assumptions the VAR estimate $\hat{\theta}$ for the reduced-form parameter θ satisfies $\sqrt{T}\left(\hat{\theta} - \theta\right) \to_d N(0, \Sigma)$, where the asymptotic covariance matrix Σ is consistently estimable, so from this perspective θ is strongly identified. MM propose testing the hypothesis $H_0 : (\nu, \varrho) = (\nu_0, \varrho_0)$ using the statistic

$$MM = T \cdot f\left(\hat{\theta}, \nu_0, \varrho_0\right)' \hat{\Sigma}_f^{-1} f\left(\hat{\theta}, \nu_0, \varrho_0\right),$$

where $\hat{\Sigma}_f$ is a Δ-method estimator for the asymptotic variance of $f\left(\hat{\theta}, \nu_0, \varrho_0\right)$, together with χ^2_6 critical values.[5] MM argue that this test controls size regardless of whether the parameters ν and ϱ are weakly identified.

The minimum-distance approach studied in this paper provides an alternative route to inference on (ν, ϱ). For any fixed (ν_0, ϱ_0) the restriction $f(\theta, \nu_0, \varrho_0) = 0$ defines a 6-dimensional manifold in the 12-dimensional space of reduced-form parameter values θ, $S_{\nu_0,\varrho_0} = \{\theta \in \mathbb{R}^{12} : f(\theta, \nu_0, \varrho_0) = 0\}$. Hence, to test $H_0 : (\nu, \varrho) = (\nu_0, \varrho_0)$ we can consider

[5] See the Supplementary Appendix for details. MM propose other identification-robust tests as well, but since these tests produce similar confidence sets in their empirical application we will focus on this one, which is the most closely related to our MD test.

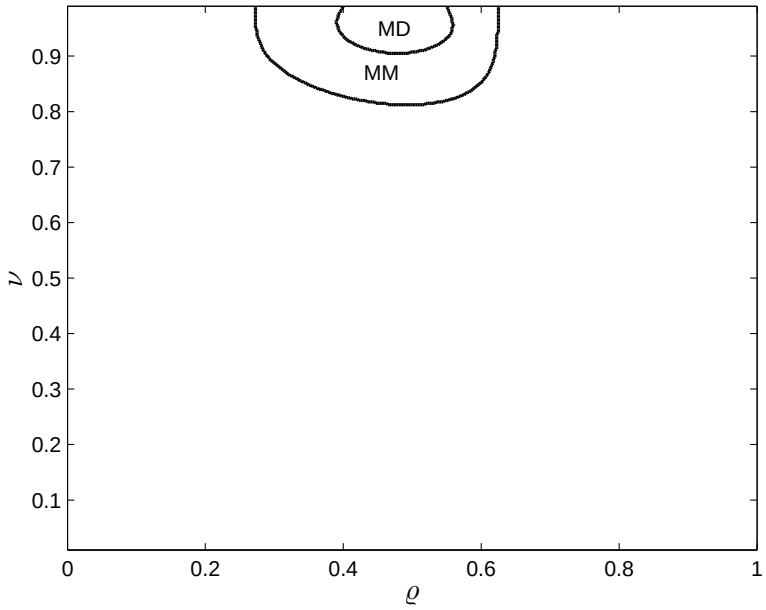


Figure 4: 95% confidence sets based on inverting our robust MD test and the MM test proposed by Magnusson and Mavroeidis (2010). Robust MD test compares the MD statistic $MD_{\nu,\varrho}$ to χ^2_8 critical values, while the MM test compares the MM statistic to χ^2_6 critical values.

the statistic

$$MD\left(\nu_0,\varrho_0\right)=\min_{\theta\in S_{\nu_0,\varrho_0}} T\cdot\left(\hat{\theta}-\theta\right)'\hat{\Sigma}_\theta^{-1}\left(\hat{\theta}-\theta\right)$$

together with robust critical values based on the curvature of the manifold S_{ν_0,ϱ_0}.

6.2 Empirical Results

We calculate 95% confidence sets for ν and ϱ based on the MD statistic together with our robust critical values and compare these confidence sets to those obtained from inverting the MM test. We use the same data as MM, which consists of quarterly observations on US GDP deflator (π_t) and labor share (x_t) from 1984 to 2008- see MM for details.

To implement our robust MD test, we need to calculate an appropriate curvature-based critical value. We show in the Supplementary Appendix that for all (ν_0,ϱ_0) the manifold S_{ν_0,ϱ_0} can be parametrized by a 6-dimensional sub-vector of VAR parameters θ_1, $\theta=\theta\left(\theta_1,\nu_0,\varrho_0\right)\in S_{\nu_0,\varrho_0}$. We find that the maximal curvature of S_{ν_0,ϱ_0} is quite high for many values (ν_0,ϱ_0), but that by projecting over two elements of θ_1 we can reduce curvature with respect to the remaining parameters to zero. Hence, for all values (ν_0,ϱ_0) we use robust critical values equal to the 95th percentile of a χ^2_8 distribution, or equivalently $F_{0.95}\left(\infty,12,4\right)$ where $p=4$ rather than 6 because we project over two

	ν	ρ
MD	$[0.91, 0.99]$	$[0.39, 0.56]$
MM	$[0.81, 0.99]$	$[0.27, 0.63]$

Table 1: 95% confidence intervals for ν and ϱ based on our robust MD test and the MM test. In both cases, confidence intervals correspond to projection-method confidence intervals obtained from joint confidence sets in Figure 4.

parameters. To calculate our robust MD confidence set for the parameters ν and ϱ jointly, we take a fine grid of values (ν_0, ϱ_0) and for each value in this grid test the hypothesis $H_0 : (\nu, \varrho) = (\nu_0, \varrho_0)$. The confidence sets consist of all those values not rejected, and are plotted in Figure 4. As we can see from this figure, in this application our robust MD confidence set is contained entirely within MM's confidence set.

In addition to building joint confidence sets for (ν, ϱ) together we can construct confidence intervals for these parameters separately. To do so, we first check whether it is possible to obtain smaller critical values by calculating curvature with respect to ν or ϱ. We find that curvature with respect to these parameters is quite high and that the smallest robust critical value we can obtain again corresponds to a χ^2_8 distribution. Thus, our robust MD confidence intervals for ν and ϱ correspond to the projection of the joint MD confidence set on the coordinate axes, and are reported in Table 1. From these results, we again see that confidence intervals based on our robust MD test are significantly smaller than those based on the MM test in this context. One reason the confidence sets are small is that the NKPC model is barely compatible with the estimated VAR coefficients. Indeed, a test of overidentifying restrictions using our robust critical values rejects the model at the 10%, though not 5%, level.

6.3 Size Simulation

One might worry that, despite the asymptotic validity of our robust MD test, the small volume of robust MD confidence sets in this application might reflect poor size control for the model and sample size considered. To address this issue we calculate the size of our robust MD test and the MM test for $H_0 : (\nu, \varrho) = (\nu_0, \varrho_0)$ in a model calibrated to match the empirical application.[6] For comparison, we also calculate the size of the non-robust MD test that incorrectly treats all parameters as strongly identified, which uses χ^2_6 critical values, and the projection-method MD test, which uses χ^2_{12} critical values. The simulated

[6]For details on the simulation design, see the Supplementary Appendix.

	Robust MD	Non-robust MD	Projection MD	MM
Size	7.0%	13.5%	1.9%	11.9%

Table 2: Size of nominal 5% tests of $H_0 : (\nu, \varrho) = (\nu_0, \varrho_0)$ based on 100 observations. Robust MD compares the MD statistic to χ^2_8 critical values (based on the curvature), while non-robust MD and projection MD use χ^2_6 and χ^2_{12} critical values, respectively. MM is the AR-MD test proposed in Magnusson and Mavroeidis (2010). Size is calculated based on 10,000 simulations.

size of these tests is reported in Table 2. As might be expected, the non-robust MD test substantially over-rejects, while the projection-method MD test is conservative. With simulated size 7% our robust MD test is only mildly over-sized, reflecting small-sample issues in the distribution of $\hat{\theta}$ and $\hat{\Sigma}_\theta$. Finally, we see that the MM test substantially over-rejects. While this may seem surprising given that Magnusson and Mavroeidis show that this test is robust to identification issues for ν and ϱ, the χ^2_6 distribution of MM under the null relies on the Δ-method approximation $f\left(\hat{\theta}, \nu_0, \varrho_0\right) \approx \frac{\partial}{\partial \theta} f\left(\theta, \nu_0, \varrho_0\right)\left(\hat{\theta} - \theta\right)$. Such an approximation, if accurate, would also suggest low curvature for the manifold S_{ν_0,ϱ_0} which is inconsistent with the data. As a result, in the sample-size considered the Δ-method approximation underlying the MM test is unreliable, while our robust MD test successfully accounts for the curvature of S_{ν_0,ϱ_0}. Hence, in addition to yielding smaller confidence sets in this application, robust MD confidence sets also have better finite-sample coverage than MM confidence sets for the parameter values considered.

7 Simulation Example: DSGE Model

In this section we further illustrate our results by considering tests of Dynamic Stochastic General Equilibrium (DSGE) model parameters in a small simulation example. DSGE models are highly non-linear, very multi-dimensional dynamic models that describe the evolution of the main macro indicators in the economy. These models are currently quite popular in applied Macroeconomics and are used by many central banks. A recent literature raises concerns about identification in these models (e.g. Ruge-Murcia (2007), Canova and Sala (2009), Iskrev (2010)), and many authors have noted that standard frequentist statistical procedures seem to be unreliable. The source and extent of weak identification in these models is not well understood, and it is typically impossible to distinguish which parameters are weakly identified using currently-available procedures. Motivated by these issues, several recent papers (Dufour, Khalaf and Kichian (2009), Guerron-Quintana, Inoue and Kilian (2013), I. Andrews and Mikusheva (2013), and

Qu (2013)) suggest tests for full parameter vector hypotheses which are robust to weak identification. With the exception of I. Andrews and Mikusheva (2013) and Guerron-Quintana, Inoue and Kilian (2013), these papers use the projection method for inference on subsets of parameters which, due to the high dimension of the parameter vector, tends to be quite conservative.

Most DSGE models can be cast into our framework, which is natural here as one suggestion for how to estimate DSGE models is through two-step matching procedures (Christiano and Eichenbaum (1992), Rotemberg and Woodford (1997), Ruge-Murcia (2012), Andreasen, Fernandez-Villaverde, and Rubio-Ramirez (2013)). Log-linearized DSGE models can be written in state-space form

$$z_t = A(\widetilde{\beta})z_{t-1} + B(\widetilde{\beta})u_t,$$

where z_t is a vector of state-space variables at time t, u_t are i.i.d. mean zero shocks with identity covariance matrix, the state-space parameter matrices $A(\widetilde{\beta})$ and $B(\widetilde{\beta})$ are non-linear functions of the structural parameter $\widetilde{\beta}$ which typically need to be evaluated numerically, and we observe $x_t = Cz_t$. In this context, a natural choice of reduced-form parameters is the auto-covariances of the observed vector-series x_t. In particular, let $\Sigma_x(j)$ be j-th order auto-covariance of x_t (for details see Iskrev (2010)):

$$\theta_j(\widetilde{\beta}) = vec(\Sigma_x(j)) = (C \otimes CA^j)(I - (A \otimes A))^{-1}vec(BB').$$

One may choose the reduced form parameter θ to be some subset of $vec(\Sigma_x(j))$, so $\theta = m(\widetilde{\beta}) = W(\theta_0(\widetilde{\beta})', ..., \theta_j(\widetilde{\beta})')'$, where W is a selection matrix. In the absence of persistence (exact or near unit roots) the usual estimators

$$\hat{\theta}_j = vec\left(\frac{1}{T-j-1}\sum_{t=1}^{T-j}(x_{t+j} - \overline{x})(x_t - \overline{x})'\right)$$

of θ_j satisfy a central limit theorem and achieve normality quite quickly. As a result, normal approximations to the distribution of $\hat{\theta} = W(\hat{\theta}_0, ..., \hat{\theta}_j)$ are usually reliable for realistic sample sizes. Hence, we can conduct inference on the structural parameters β using a minimum-distance approach focused on matching auto-covariances.

While our discussion focuses on log-linearized DSGE model, one could equally well

apply our approach to nonlinear DSGE models, for example matching covariances calculated from higher-order approximations as in Andreasen, Fernandez-Villaverde, and Rubio-Ramirez (2013). Even in the log-linear case we could base inference on quantities other than auto-covariances, for example matching the parameters of the state-space representation. Indeed, by matching the parameters of the matrices $A(\widetilde{\beta})$ and $B(\widetilde{\beta})$ we may be able to obtain a more powerful test. However, one should be cautious with this choice of reduced-form parameters since it is known that state-space coefficients are sometimes poorly identified, casting doubt on the normal approximation to the distribution of state-space parameter estimates. In particular, if the dynamics of the data are described by a VARMA process such that some VAR and MA roots are close to each other (near-canceling roots), the corresponding state-space representation coefficients are weakly identified. This issue has been raised by Schorfheide (2013) in the DSGE context and is studied in an ARMA model by D. Andrews and Cheng (2012).

7.1 Simulation study: A Small-scale DSGE Model

We apply our approach to a small-scale DSGE model based on Clarida, Gali and Gertler (1999). The (log-linearized) equilibrium conditions for the model are

$$\begin{cases} bE_t\pi_{t+1} + \kappa x_t - \pi_t + \varepsilon_t = 0, \\ -[r_t - E_t\pi_{t+1} - rr_t^*] + E_t x_{t+1} - x_t = 0, \\ \lambda r_{t-1} + (1-\lambda)\phi_\pi \pi_t + (1-\lambda)\phi_x x_t + u_t = r_t, \\ rr_t^* = \rho \Delta a_t, \end{cases} \tag{15}$$

where the exogenous variables (Δa_t and u_t) evolve according to

$$\Delta a_t = \rho \Delta a_{t-1} + \varepsilon_{a,t}; \quad u_t = \delta u_{t-1} + \varepsilon_{u,t};$$

$$(\varepsilon_t, \varepsilon_{a,t}, \varepsilon_{u,t})' \sim iid\, N(0, \Sigma); \Sigma = diag(\sigma^2, \sigma_a^2, \sigma_u^2).$$

Here we assume that a researcher observes data on inflation π_t, the interest rate r_t and some measure of real activity x_t. This model has ten parameters: the discount rate b, the structural parameters κ, ϕ_x, ϕ_π, and λ, and the parameters describing the evolution of the exogenous variables. We set $b = 0.99$ and estimate the remaining parameters by maximum likelihood using de-meaned US macro data from Smets and Wouters (2007). Since the estimated values of ρ and κ are quite close to the boundaries of their respective parameter spaces (one and zero respectively), we set $\rho = .85$ and $\kappa = .1$ while calibrating the remaining parameters to their ML estimates, in particular taking $\phi_x = 2.278$, $\phi_\pi =$

2.023, $\lambda = .898$, $\delta = .103$, $\sigma_a = .325$, $\sigma_a = .265$, and $\sigma = .556$. We generate samples of size 300 from this model and then discard the first 100 observations, using only the last 200 observations for the remainder of the analysis. Given well-documented problems with estimating b in many models, from this point forward we treat this parameter as known and focus on the remaining nine parameters.

As discussed above, to test hypotheses on the structural parameters in DSGE models we can test the implied restriction on the auto-covariances $\theta(\widetilde{\beta})$. In particular, we let θ consist of the unique entries of the covariance matrix of (x_t, π_t, r_t) and their first auto-covariance, giving us 15 reduced-form parameters. To focus on the problem of weak identification and abstract from the issues which may arise from HAC covariance matrix estimation, we treat the true covariance matrix Σ of our reduced-form parameter estimates as known.

We consider the problem of separately testing that each of the structural parameters is equal to its true value (as one needs to do to construct confidence sets for each parameter individually). For example, to test $H_0 : \kappa = \kappa_0$, we let β contain all the parameters other than κ (that is, $\widetilde{\beta} = (\kappa, \beta)$) and consider the MD statistic $MD(\kappa_0) = \min_\beta T(\widehat{\theta} - \theta(\kappa_0, \beta))'\Sigma^{-1}(\widehat{\theta} - \theta(\kappa_0, \beta))$. As before, the key issue is what critical values to use. The projection method uses the 95th percentile of a χ^2_{15}, which is equal to 25. If we assume that a p-dimensional sub-vector of β is strongly identified, we can use χ^2_{15-p} critical values instead, which are equal to 14.07 (for $p = 8$). Applying our robust critical values, in contrast, requires no assumption on the strength of identification. As we might expect in a poorly identified model, tests which concentrate out the nuisance parameters do not have correct size. We simulated tests for each parameter separately, and all of them over-reject, though the degree of over-rejection is limited. The largest size distortion occurs for tests of $\kappa = \kappa_0$, where nominal 5% tests have true size 11.4%.

For each of the nine parameters we calculate the curvature of the submanifold of $\{\Sigma^{-\frac{1}{2}}\theta(\beta)\}$ obtained by holding that parameter equal to its null value, intersected with the ball of radius $\sqrt{2}R$ around θ_0, for R the .99 quantile of a χ^2_{15} distribution. We find quite substantial curvature: the manifold corresponding to $\kappa = \kappa_0$, for example, has a maximal curvature of 5.16, which if plugged into $F(C, R, 15, 8)$ implies a robust critical value nearly as large as that used by the projection method.

These large values of curvature result from calculating curvature with respect to all nuisance parameters simultaneously. As noted above, however, if some nuisance pa-

Parameter tested	Robust Critical Value	Robust Test Size	Projection Size
ϕ_x	17.3	2.5%	0.4%
ϕ_π	17.3	2.8%	0.6%
λ	20.4	1.0%	0.3%
ρ	20.2	1.4%	0.6%
δ	20.2	0.9%	0.3%
κ	22.5	2.4%	1.7%
σ_a	17.7	3.0%	0.4%
σ_u	19.0	2.4%	0.6%
σ	19.5	1.5%	0.6%

Table 3: Nominal 5% Tests of one-dimensional hypotheses on structural parameters. The first column lists the tested parameter for each row, while the other parameters are treated as nuisance parameters. The statistic is AR minimized over nuisance parameters. Projection method critical values are 25. Robust critical values are obtained by considering projection over ϕ_x, ϕ_π, (ϕ_x, ϕ_π), $(\phi_x, \phi_\pi, \lambda)$, (ϕ_x, ϕ_π, ρ), $(\phi_x, \phi_\pi, \kappa)$ and using the smallest critical value

rameters correspond to directions of high curvature (we may loosely call them weakly identified) while others enter nearly linearly, we may be able to substantially reduce our critical values by projecting over the weakly identified parameters. Specifically, we find that projecting over ϕ_x and ϕ_π often substantially reduces curvature with respect to the remaining parameters. Motivated by this fact we calculate curvature projecting over the following subsets of parameters: ϕ_x, ϕ_π, (ϕ_x, ϕ_π), $(\phi_x, \phi_\pi, \lambda)$, (ϕ_x, ϕ_π, ρ), $(\phi_x, \phi_\pi, \kappa)$. To test hypotheses on each of the nine structural parameters we use the smallest robust critical value obtained by projecting over one of these six subsets.

For each structural parameter, Table 3 reports the robust critical value obtained from this exercise (column 2), together with the simulated size (based on 10,000 simulations) of nominal 5% tests based on our robust critical values (column 3), and projection-method tests (column 4). As we can see, projection-method based tests (using critical values of 25) are typically quite conservative, with simulated size less than or equal to 0.6% for all parameters but κ. In contrast, for most parameters our robust critical values fall about halfway between projection method critical values (25) and non-robust critical values (14.07). Tests based on our robust critical values control size and, while conservative, are for the most part substantially less so than projection-method tests.

8 Appendix with proofs

8.1 Proof of Theorem 1

The proof is based on the following lemma:

Lemma 2 *Assume the curve* $\alpha(s) : [0, b] \to D_C \subset \mathbb{R}^k$ *is parameterized by arc length and that its curvature* $\kappa(s) = \|\ddot{\alpha}(s)\| \leq \frac{1}{C}$ *for all points* s. *Assume that* $\alpha(0) = 0$ *and* $\dot{\alpha}(0) = v \in$ *span*$\{e_1, ..., e_p\}$, *where* $e_1, ..., e_p$ *are first* p *basis vectors. Then the curve* $\alpha(s)$ *is contained in the set* $M_v \cap D_C$, *where*

$$M_v = \{x : \langle x, v\rangle^2 + (C - \|x - \langle x, v\rangle v\|)^2 \geq C^2\}. \tag{16}$$

Proof of Lemma 2.

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

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

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

$$\langle v, \beta(t)\rangle = \langle v, \dot{\alpha}(t)\rangle \geq \cos(\frac{t}{C}). \tag{17}$$

Since $\alpha(s)$ is parameterized by arc length, from inequality (17) we have:

$$\|\dot{\alpha}(t) - \langle v, \dot{\alpha}(t)\rangle v\| \leq |\sin(\frac{t}{C})|. \tag{18}$$

This, in turn, implies that

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

Inequality (17) also implies that

$$\langle v, \alpha(t)\rangle \geq \int_0^t \cos(\frac{s}{C})ds = C\sin(\frac{t}{C}). \tag{19}$$

Combing these results yields

$$\langle v, \alpha(t)\rangle^2 + (C - \|\alpha(t) - \langle v, \alpha(t)\rangle v\|)^2 \geq C^2$$

for all $t \leq \frac{\pi}{2}C$. Notice that (19) implies that for $\tau = \frac{\pi}{2}C$ we have $\langle v, \alpha(\tau)\rangle \geq C$ and thus for the first p coordinates of $\alpha(\tau)$, which we denote $\alpha^{(1)}(\tau)$, we have $\|\alpha^{(1)}(\tau)\| \geq C$ so the curve is leaving or has already left the cylinder D_C and thus $b \leq \frac{\pi}{2}C$. This concludes the proof of the lemma. □

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

$$\bigcup_{\substack{v\in T_0(S)\\ \|v\|=1}} M_v = \{\|x^{(1)}\|^2 + (C - \|x^{(2)}\|)^2 \geq C^2\} = \mathcal{M}, \tag{20}$$

where M_v is defined in (16), $\mathcal{M}$ is defined in (9) 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

$$\langle x, v\rangle^2 + (C - \|x - \langle x, v\rangle v\|)^2 \geq C^2. \tag{21}$$

For each x let us find the maximum of the expression on left-hand side of inequality (21) over $v \in T_0(S), \|v\| = 1$:

$$\begin{aligned} &\langle x, v\rangle^2 + (C - \|x - \langle x, v\rangle v\|)^2 = \\ &= \langle x, v\rangle^2 + C^2 + \|x\|^2 - \langle x, v\rangle^2 - 2C\|x - \langle x, v\rangle v\| = \\ &= C^2 + \|x\|^2 - 2C\|x - \langle x, v\rangle v\| \end{aligned}$$

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 (21) 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 (21) equals

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

This proves statement (20).

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) $\alpha(s)$ connecting q and 0 lying in S_C, where such a curve exists since S_C is a connected manifold. Parameterize this curve by the arc length. The curve $\alpha(s)$ is a geodesic in S if and only if at any point $q = \alpha(t)$ the second derivative $\ddot{\alpha}(t)$ 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 α at each point $q = \alpha(t)$ is equal to $\kappa_q(X, S)$ (where $X = \dot{\alpha}(t)$), 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 2 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, $\widetilde{u}$ 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 will put forward two statements: first, that all points in the intersection of $R_\tau \bigcap \mathcal{M} \bigcap D_C$ are not 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 $S_C \subseteq \mathcal{M} \bigcap D_C$. $\square$

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 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 $\sqrt{\chi^2_p}$ and $\sqrt{\chi^2_{k-p}}$ distributions, respectively. $\square$

8.2 Proof of Lemma 1

Proof of Lemma 1. 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 C be the largest value for which all assumptions of Theorem 1 are satisfied. Let $\psi_C(\xi, R)$ be defined as

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

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 Section 4.1, but is defined on a different probability space, as $\psi_C(R)$ is written in terms of 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 $\psi_C(\xi, R) \ge F_{1-\alpha}(C, R, k, p)$. The size $E\varphi(\xi) = \alpha$, so since $P\{\chi^2_k \ge R^2\} < \alpha$ we know that φ rejects for all realizations of ξ where $\|\xi\| > R$ as $\|\xi\| \ge \rho(\xi, N_{\widetilde{u}})$. 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 1 is

$$\widetilde{\varphi} = \begin{cases} 1, & \text{if } MD \ge F_{1-\alpha}(C^*_R, R, k, p); \\ 0, & \text{otherwise.} \end{cases} \tag{22}$$

We claim that $\widetilde{\varphi} \le \varphi$ 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, $\|\xi\| \le R$, then the cylinder $\widetilde{D}_R(x_0)$ around $x_0 = \Sigma^{-1/2}\theta_0$ lies inside of ball B^* of radius $(1 + \sqrt{2})R$ around $\hat{x} = x_0 + \xi$, and thus

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

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

Note that the function $F_{1-\alpha}(c, R, k, p)$ is decreasing in c, and hence $F_{1-\alpha}(C, R, k, p) \le F_{1-\alpha}(C^*_R, R, k, p)$. Further, all the assumptions of Theorem 1 are satisfied so $MD =$

$\rho^2(\xi, S) \leq \rho^2(\xi, N_{\widetilde{u}}) \leq \psi_C(\xi, R)$. Combining these results we obtain that

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

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. □

9 References

Anderson, T., and H. Rubin (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20, 46-63.

Andreasen, M. M., J. Fernandez-Villaverde, and J. F. Rubio-Ramirez (2013): "The Pruned State-Space System for Non-Linear DSGE Models: Theory and Empirical Applications," *unpublished manuscript.*

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 (2013): "Maximum Likelihood Inference in Weakly Identified Models," *unpublished manuscript.*

Canova, F., and L. Sala (2009): "Back to Square One: Identification Issues in DSGE Models," *Journal of Monetary Economics*, 56, 431-449.

Christiano L.J. and M. Eichenbaum (1992): "Current Real-Business-Cycle Theories and Aggregate Labor-Market Fluctuations," *American Economic Review*, 82, 430-450.

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

Dufour, J.M., (2003): "Identification, Weak Instruments, and Statistical Inference in Econometrics," *The Canadian Journal of Economics*, 36, 767-808.

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., L. Khalaf, and M. Kichian (2006): "Inflation Dynamics and the New Keynesian Phillips Curve: An Identification Robust Econometric Analysis," *Journal of Economic Dynamic and Control*, 30, 1707-1727.

Dufour, J.M., L. Khalaf, and M. Kichian (2009): "Structural Multi-Equation Macroeconomic Models: Identification-Robust Estimation and Fit," *Bank of Canada Working Paper.*

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.

Guerron-Quintana, P., A. Inoue, and L. Kilian (2013): "Frequentist Inference in Weakly Identified DSGE Models," *Quantitative Economics*, 4(2), 197-229.

Guggenberger, P. and R.J. Smith (2005): "Generalized Empirical Likelihood Estimators and Tests under Partial, Weak and Strong Identification," *Econometric Theory*, 21, 667-709.

Guggenberger, P. and R.J. Smith (2008): "Generalized Empirical Likelihood Tests in Time Series Models with Potential Identification Failure," *Journal of Econometrics*, 142(1), 134-161.

Inoue, A. and B. Rossi (2011): "Testing for Weak Identification in Possibly Nonlinear Models," *Journal of Econometrics*, 161(2), 246-261.

Iskrev, N. (2010): "Local Identification in DSGE," *Journal of Monetary Economics*, 57(2), 189-202.

Kleibergen, F. (2005): "Testing Parameters in GMM Without Assuming that They Are Identified," *Econometrica*, 73, 1103-1123.

Kleibergen, F. (2007): "Generalizing Weak Instrument Robust IV Statistics towards Multiple Parameters, Unrestricted Covariance Matrices and Identification Statistics," *Journal of Econometrics*, 139(1), 181-216.

Kleibergen, F., and S. Mavroeidis (2009): "Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve," *Journal of Business & Economic Statistics*, 27(3), 293-311.

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

Magnusson L.M. and S. Mavroeidis (2010): "Identification-Robust Minimum Distance Estimation of the New Keynesian Phillips Curve," *Journal of Money, Banking, and Credit*, 42, 465-481.

Mavroeidis, S. (2005): "Identification Issues in Forward-looking Models Estimated by GMM with an Application to the Phillips Curve," *Journal of Money, Credit and Banking*, 37, 421-449.

Mavroeidis, S. (2010): "Monetary Policy Rules and Macroeconomic Stability: Some New Evidence," *American Economic Review*, 100(1), 491-503.

Mavroeidis S., M. Plagborg-Moller, J. Stock (2013): "Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve," Unpublished Manuscript.

Nason, J.M. and G.W. Smith (2008): "Identifying the New Keynesian Phillips Curve," *Journal of Applied Econometrics*, 23, 525-551.

Qu, Z. (2013): "Inference and Specification Testing in DSGE Models with Possible Weak Identification," *Quantitative Economics, forthcoming.*

Rotemberg, J., and M. Woodford (1997): "An Optimization Based Econometric Framework for the Evaluation of Monetary Policy," *NBER Macroeconomic Annual*, 12, 297-346.

Ruge-Murcia, F.J. (2007): "Methods to Estimate Dynamic Stochastic General Equilibrium Models," *Journal of Economic Dynamics & Control*, 31, 2599-2636.

Ruge-Murcia, F.J. (2012): "Estimating Nonlinear DSGE Models by the Simulated Method of Moments: With an Application to Business Cycles," *Journal of Economic Dynamics and Control*, 36 (6), 914-938.

Sbordone, A. M. (2005), "Do expected future marginal costs drive inflation dynamics?" *Journal of Monetary Economics*, 52, 1183-1197.

Schorfheide, F. (2013): "Estimation and Evaluation of DSGE Models: Progress and Challenges," Forthcoming in: D. Acemoglu, M. Arrelano, and E. Deckel (eds.): *Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress*, Cambridge University Press.

Smets, F. and R. Wouters (2007): "GMM with Weak Identifiication," *Econometrica*, 68, 1055-1096.

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.

Stock, J.H., J.H. Wright, and M. Yogo (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business & Economic Statistics*, 20(4), 518-529.

Stock, J.H. and M. Yogo (2005): "Testing for Weak Instruments in Linear IV Regression," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg,* D. W. K. Andrews and J. H. Stock, eds., Cambridge: Cambridge University

Press, pp. 80-108.

Wright, J.H. (2003): "Detecting Lack of Identification in GMM," *Econometric Theory*, 19, 322-330.

Yogo, M., (2004): "Estimating the Elasticity of Intertemporal Substitution when Instruments are Weak," *Review of Economics and Statistics* 86, 797-810.